REAL-TIME PRODUCTION AND SETUP SCHEDULING
OF DETERMINISTIC AND STOCHASTIC
MANUFACTURING SYSTEMS

By

MOHSEN EL HAFSI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

This dissertation is dedicated to

my parents, Bou Ali and Najet,

my loving wife Essia,

and my beautiful son, Mohamed Amine.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my deepest thanks and sincere appreciation to Dr. Boghos Sivazlian, chairman of my supervisory committee, for all his invaluable advice and guidance at both the professional and the social levels.

I wish to acknowledge and express my appreciation to Dr. Sherman Bai, who served as cochairman of my supervisory committee and immediate supervisor of my dissertation, for all his invaluable advice, guidance and encouragement throughout this research. His guidance and instruction helped the shape and form of this research, as well as, my personal and career development.

I would also like to thank the other members of my committee, Dr. Jack Elzinga, Dr. Chung Lee and Dr. Mark Yang, for their comments and suggestions concerning this research.

Thanks are also extended to the Department of Industrial and Systems Engineering at the University of Florida for providing financial support.

Finally, I wish to express the most heartfelt gratitude to my dearest wife Essia for her constant support and encouragement throughout my studies. And, to my parents Bou Ali and Najet for all the sacrifices they endured and continue to endure so that I succeed and excel in life.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

REAL-TIME PRODUCTION AND SETUP SCHEDULING
OF DETERMINISTIC AND STOCHASTIC
MANUFACTURING SYSTEMS

By

Mohsen El Hafsi

August, 1995

Chairman: Boghos D. Sivazlian
Cochairman: Sherman X. Bai
Major Department: Industrial and Systems Engineering

Many manufacturing systems incur a significant setup time and/or
setup cost whenever the production resource is set up for a new product after it
has been used for another. Setups require a downtime during which there is no
production; in turn, this might result in decreased production. Cost may be the
result of material consumption due to testing (scraps), energy consumption,
man hours lost, and tolerance adjustment of the machine for the next part.
This situation leads to the following decision making process: when to stop this
multi-product system from manufacturing the current product or part, and
when to start manufacturing the next product. Basically, the problem is to
select a batch size for each product to be manufactured and to sequence the
products on the machine.

The purpose of this research is to provide real-time optimal or near-
optimal feedback policies for determining production rates and setup times for
deterministic and stochastic systems (i.e., machine subject to failures and
repairs), where the demand of each product is assumed to occur at a constant

rate. The most critical issue in real-time scheduling and planning of manufacturing systems is the computational effort required to achieve good performance. To reduce the complexity of the system, most of the quantities of interest are modeled as continuous. We deal with continuous parameters, such as production and demand rates and cumulative production surplus (inventory or backlog).

In order to gain insight and develop solution techniques, we first deal with a two-product system. That is, the real-time production and setup scheduling of a one-machine two-product manufacturing system. The machine can be either perfectly reliable (deterministic system) or unreliable (stochastic system).

The deterministic system is formulated as a feedback control problem, where the state of the system is measured by the amount of cumulative production surplus (inventory if positive and backlog if negative). The performance measure or objective is to minimize the total setup, inventory and backlog costs over a finite planning horizon. The optimal steady state and transient solutions are obtained separately.

The stochastic system is formulated as a stochastic control problem. The objective is to minimize the expected total discounted cost of setup, inventory and backlog over an infinite planning horizon. To obtain the optimal policy, a numerical approach is adopted. Based on the results of the deterministic model, a simple heuristic policy, implementable in real-time, is proposed. Simulation results show that the heuristic policy performs very well when applied to manufacturing systems involving low failure rates.

CHAPTER 1
INTRODUCTION

1.1 Motivation

Many manufacturing systems can produce more than one kind of product. In most, a significant cost and/or time is incurred whenever the manufacturing system is set up to produce a new product after it has been used for another. Setups require a down time during which there is no production, which in turn, results in decreased production. Cost may be the result of material consumption, due to testing and tolerance adjustment of the machine for the next part. The reader is referred to the book by Gershwin (1993) for good examples of manufacturing systems involving setup time and/or cost.

There are situations where it could be more economical to purchase one machine that is capable of processing many products or parts than to purchase many dedicated machines. This situation leads to the question of how to decide when to stop this multiple-product machine from producing the current product or part, and which product to start making next. Basically, the issue is to select both a sequence in which products will be produced and a batch size for each product to be produced. If setups are performed too often, high change over costs are incurred; and, if too much time is spent on setups, there is not enough time to produce the required amount of the product to meet the demand. On the other hand, if setups are not performed often enough, larger

1

inventory levels of the part being produced may result while shortages may occur for the parts not being produced.

The major goal of this research is to provide real-time optimal or near-optimal feedback policies for deciding on production rates and setup times, for both stochastic (subject to failures and repairs) and deterministic systems, in which one is only concerned with operations and setups. In order to gain insight and develop solution techniques, we will be dealing only with small scale systems. However, although small scale, the problems under consideration are still very challenging and practically meaningful. The most critical issue in real-time scheduling and planning of manufacturing systems is the computational effort required to achieve good performance. To reduce the complexity of the system, most of the quantities of interest will be modeled as continuous. Basically, we will be dealing with continuous parameters, such as production and demand rates and production surplus. Once the solution is obtained, it is then easy to convert the production rate into discrete loading times at the machine.

In the remainder of this introductory chapter, we will review some models of setup planning without feedback then give an overview of the literature on real-time setup scheduling with feedback.

1.2 Production and Setup Planning Models

There has been an enormous amount of literature on the subject of setups, lot sizes, and related issues. We do not survey all of it; we mostly review the part of the literature that is related to the kind of problems we are dealing with in this research proposal. The models described in this chapter are chosen because of either historical or economical importance. Furthermore,

these models form a good starting point for the real-time scheduling models presented in this proposal. In section 1.2.1 we present one of the oldest and simplest lot sizing models, the *Economic Order Quantity* model. In section 1.2.2, we present an elegant extension of this model due to Wagner and Whitin (the Wagner-Whitin model). In section 1.2.3, the Wagner-Whitin (1958) formulation is extended for the case of multiple parts under capacity restrictions (Newson's model, 1975a, 1975b). In section 1.2.4, the *Economic Lot Scheduling Problem*, which is a more detailed model, is presented.

1.2.1 The economic order quantity model

The simplest model of lot sizing is the Economic Order Quantity or EOQ model. The EOQ model does not describe a true setup manufacturing system because it deals with only a single item or product. It describes a system that acquires items rather than produces them. The items are replenished at discrete times rather than continuously due to ordering costs rather than setup costs. There are many variants of the EOQ formula. A simple version is given by

$$Q = \sqrt{\frac{2KD}{h}} \tag{1.1}$$

where Q is the economic order quantity, K is a fixed ordering cost incurred each time an order is placed, D is the constant demand rate of the items and h represents the dollar cost per unit of holding an item in inventory.

Although this is a very simplified model, the EOQ formula describes a feedback control policy. First, given the quantities K, D, and h, we compute Q. Using Q, the inventory level is monitored as follows: whenever it reaches 0, order Q. Here, the inventory level represents the state of the system. Hence,

this is a feedback, or real-time, control policy since the ordering action is based on a measurement of the system state.

Other variations of the EOQ model have been studied. These include the effect of lead time, the effect of random demand and the effect of shortage, to cite a few.

<u>1.2.2 Extension of the EOQ model: The Wagner-Whitin model</u>

Wagner and Whitin (1958) extend the EOQ model by dropping the assumption of a steady-state demand rate. They divide the time into periods of fixed length. The demand for items is known during each period. The model assumes a single product. Their objective is to minimize the total inventory holding costs. They treat material as a continuous quantity.

The dynamics of the system is described by the following difference equation:

$$I_t = I_0 + \sum_{i=1}^{t-1} x_i - \sum_{i=1}^{t-1} d_i \geq 0$$

where I_t is the inventory entering period t (state variable), x_t is the amount of material procured or produced during period t (decision or control variable), and d_t is the amount demanded during period t.

The problem is formulated as a dynamic programming problem with value function f_t satisfying

$$f_t(I_t) = \min_{\substack{x_t \geq 0 \\ I_t + x_t \geq 0}} \left[i_{t-1} I_t + \delta(x_t) s_t + f_{t+1}(I_t + x_t - d_t) \right], \; t = 1, \ldots, N;$$

$$f_N(I_N) = \min_{\substack{x_N \geq 0 \\ I_N + x_N = d_N}} \left[i_{N-1} I_N + \delta(x_N) s_N \right]$$

where i_t is the interest charge per unit of inventory carried forward to period $t+1$, s_t is the ordering or setup cost, N is the number of periods in the planning horizon and $\delta(x_t)$ is a binary function given by

$$\delta(x) = \begin{cases} 0 & \text{if } x = 0; \\ 1 & \text{if } x > 0. \end{cases}$$

Notice that this model assumes that the resource has to be set up if an item is manufactured during time period t regardless of whether the same kind of item was manufactured during $t-1$. The optimal solution of this problem suggests that new production occurs only when the inventory becomes zero. The amount produced each time is exactly enough to meet the demand until the next time the inventory goes to zero (this makes sense because the system is deterministic and the demand is known in advance).

Because the Wagner-Whitin model assumes perfect knowledge of the demand for future periods (deterministic system), it is appropriate for production planning rather than real-time scheduling.

1.2.3 Extensions of the Wagner-Whitin model

A significant extension of the Wagner-Whitin formulation is due to Newson (1975a, 1975b), who explicitly represents multiple parts in the formulation as well as capacity constraints. The objective is to minimize the sum total of setup, production and inventory holding costs. Because of the capacity constraint this problem is called the *Capacitated Lot Size Problem* (CLSP).

As in the Wagner-Whitin formulation Newson divides the planning horizon into T periods of fixed length. He treats material as a continuous quantity. The problem is formulated as a Linear Programming problem. The

decision variables are I_{it}, the amount of inventory of Product i at the end of period t, and x_{it}, the amount of Product i made in period t. There are two constraints. The inventory dynamics are given by

$$I_{i,t-1} + x_{it} = I_{it} + d_{it}, \quad i = 1, \ldots, N, \; t = 1, \ldots, N;$$

$$I_{i0} \geq 0, \; i = 1, \ldots, N \text{ are given}$$

where N is the number of products to be produced and d_{it} is the demand of Product i during period t. The capacity constraint is expressed as

$$\sum_{i=1}^{N} \left[r_{ik}^{\delta} \delta(x_{it}) + r_{ik}^{x} x_{it} \right] \leq R_{kt}, \quad k = 1, \ldots, K, \; t = 1, \ldots, T$$

where r_{ik}^{δ} is the capacity lost due to a setup for Product i on resource k, r_{ik}^{x} is the per unit capacity lost due to the production of Product i on resource k, R_{kt} is the level of resource k available in period t, K is the total number of resources and $\delta(x_{it})$ has the same meaning as in the Wagner-Whitin formulation. Additional constraints include the nonnegativity of the decision variables I_{it} and x_{it} ($i = 1, \ldots, N, \; t = 1, \ldots, T$). The objective function is given by

$$f = \sum_{i=1}^{N} \sum_{t=1}^{T} s_i \delta(x_{it}) + v_i x_{it} + h_i I_{it} \,.$$

where s_i is the setup cost for Product i, v_i, the per unit production cost of Product i and h_i is the inventory holding cost per unit of Product i per period.

Like the Wagner-Whitin model, Newson's model is appropriate for production planning rather than real-time scheduling. Although this is a highly simplified formulation, there is no explicit analytic solution to this problem. To solve this problem Newson and others provide heuristics and numerical techniques.

Other extensions of the Wagner-Whitin formulation include Mixed Integer Programming formulations. These problems are known to be hard to

solve (i.e., their computational time grows exponentially with the size of the problem).

1.2.4 The economic lot scheduling problem

The Economic Lot Scheduling Problem (ELSP) is usually stated as follows: There is one or more identical machines on which m products are to be produced. The demand rate for products is assumed constant, no backlog or backorders are allowed and the system is in steady state. The problem can be viewed as one of finding lot sizes that meet the demand and minimize the average setup and inventory holding costs per unit time. Elmaghraby (1978) gives a thorough review of the models used for the ELSP. Recent research in the field includes the work of Goyal (1984), Dobson (1987), and Carreno (1990).

This section is based on Dobson's (1987) formulation. The problem is one of deciding on a cycle of length T, a production sequence F^1, ..., F^n (the sequence may contain repetitions, i.e., $n \geq m$) to be repeated indefinitely, production runs t_1, ..., t_n and idle times w_1, ..., w_n so as to meet the demand (no backlog is allowed) and minimize the sum total of inventory holding cost and setup cost.

The data for the problem are for each Product i ($i = 1$, ..., m), u_i denotes the constant production rate, d_i denotes the constant demand rate, h_i denotes the inventory holding cost, A_i denotes the setup cost and s_i denotes the setup time. Dobson's model as well as many others assumes sequence-independent setups. That is, the setup time depends only on the part to be produced next (if it is different from the previously produced).

The decision variables are the position F^i, in a given sequence where part i is produced; the cycle length T; the length of the production sequence n (since each product is allowed to be produced more than once in a production sequence, $n \geq m$); the duration of the ith production run t_i; and the idle time w_i during the ith production run.

Let $J_i = \{j | F^j = i\}$ be the set of times in the cycle, that Product i is produced. Let L_k be the set of positions in a given sequence, from k (when F^k is produced), up to the next position in the sequence that F^k is produced again. The definition of L_k assumes that the sequence $F^1, \ldots, F^n$ is circular. Let $\mathfrak{I}$ be the set of all feasible finite sequences of parts. That is, $\mathfrak{I} = \{(F^1, \ldots, F^n) | m \leq n < \infty, |J_i| \geq 1, i = 1, \ldots, m\}$. Using this notation the ELSP is formulated as follows:

$$\inf_{\substack{(F^1,\ldots,F^n) \in \mathfrak{I}}} \min_{\substack{t \geq 0 \\ w \geq 0 \\ T \geq 0}} \frac{1}{T} \sum_{j=1}^{n} \left(\frac{1}{2} h_{p^j} (u_{p^j} - d_{p^j}) \frac{u_{p^j}}{d_{p^j}} (t_{p^j})^2 + A_{p^j} \right)$$

subject to

$$\sum_{j \in J_i} u_i t_{p^j} = d_i T, \, i = 1, \ldots, m; \tag{1.2}$$

$$\sum_{j \in L_k} (t_{p^j} + s_{p^j} + w_{p^j}) = \frac{u_{p^k} t_{p^k}}{d_{p^k}}, \, k = 1, \ldots, n; \tag{1.3}$$

$$\sum_{j=1}^{n} (t_{p^j} + s_{p^j} + w_{p^j}) = T. \tag{1.4}$$

Constraints (1.2) state that the total production over a given sequence must meet the demand for each product. Constraints (1.3) state that the inventory level of Product i reaches zero just as production of Product i is about to begin. This is usually known as the Zero Switch Rule (ZSR) in the literature.

The third constraint states that the total amount of production, setup and idle times over a cycle must be the length of the cycle.

To solve this problem, Dobson develops a heuristic to determine a production sequence by first obtaining approximate production frequencies. These frequencies are rounded to powers of two. Dobson then uses a bin-packing heuristic to obtain a sequence. Additional cost penalties are incurred when the sequence is determined and the exact quantities and timing of the batches are computed.

Dobson's formulation differs from that of Newson (presented in Section 1.2.3) in that setup behavior is modeled more precisely. It is more restrictive in that the demand rate is required to be constant. Hence, Dobson's model is more appropriate for shorter planning horizon compared to that of Newson.

1.3 Real-Time Setup Scheduling with Feedback

All the models presented in the previous sections, focus on the setup scheduling problem in isolation and under steady-state conditions. They do not address nonsteady-state situations caused by disruptions such as machine failures and scheduled maintenance in a manufacturing system. Under non-steady-state conditions the problem cannot be reduced to one of lot sizing since the lot sizes will not remain constant over time. The purpose of this research is to investigate the setup scheduling problem in a feedback control framework, which could respond to various discrete events in a manufacturing system.

1.3.1 Feedback control scheduling framework

Feedback control models for production scheduling appeared in the early '80s, with the pioneering work of Kimemia and Gershwin (1983). They

developed a feedback scheduling framework and suggested that the emphasis should be on the flow rate of parts rather than on individual jobs or operations. This framework assumes manufacturing systems that are operated to meet a multiple-part demand rate as closely as possible. It focuses attention on random failures and repairs of machines affecting the production capacity. Based on the works by Rishel (1975) and Olsder and Suri (1980), Kimemia and Gershwin derived a feedback solution to the problem of dispatching parts to machines in a failure-prone Flexible Manufacturing System or FMS (an FMS is a manufacturing system that can produce many different part types with little time and expense lost in changeover from one product to another). Their approach was to separate the relatively long term issues (machine failures) from the short term decisions (part loading). They showed that the optimal control has a special structure, described by a *hedging point strategy*, in which a positive production surplus (or inventory) of parts is maintained during times when excess capacity is available to hedge against future capacity shortages brought by machine failures. The state of the system is measured by the production surplus of all parts, which is defined as the difference between the cumulative production and the cumulative demand of a part during a certain interval of time. Kimemia and Gershwin's approach is often referred to as flow control in the literature.

Recently, Gershwin (1993) proposed a more general framework suggesting a hierarchical approach for the planning and scheduling of manufacturing systems. He groups events from the least frequent (at the top level of hierarchy) to the most frequent ones (at the bottom level). If the hierarchy levels are well separated (frequency wise), each level can be formulated as a flow control problem.

1.3.2 Corridor policies

Using the formalism of Kimemia and Gershwin (1983), Sharifnia et al. (1991) studied the setup scheduling problem of a multiple part machine. They proposed Corridor Policies to time the setup switching epochs. The corridor policy does this by constructing a corridor around the production surplus trajectory in the production surplus space. The corridor is made up of a set of planes (in n-dimensional space) that are each labeled with the index of a part; when the trajectory hits a plane, the setup is changed so that the machine can produce the part corresponding to the plane just hit by the trajectory. Sharifnia et al. showed that for a round robin sequence (in which the parts are produced only once per sequence and in a cyclic fashion) and when the corridor satisfies certain conditions, the production surplus vector converges to a cyclic schedule which they refer to as the Limit Cycle.

Following the ideas of Sharifnia et al., Srivatsan and Gershwin (1990) developed methods for choosing the parameters of the corridors when the setup frequencies are not all the same.

1.3.3 Distributed real-time setup scheduling policies

Perkins and Kumar (1989) and Kumar and Seidman (1990) studied distributed deterministic systems. They proposed a class of setup scheduling policies implementable in real-time in a distributed fashion. These policies make setup change based on buffer levels upstream of a machine. To define the proposed policies, consider a system with a single machine that produces N parts. Type i parts arrive at the rate d_i at a buffer upstream of the machine. Type i parts need τ_i time units to be processed on the machine. Then, the policies are defined as follows:

   Clear the Largest Buffer (CLB): Do not change setup until the buffer of the parts being processed is cleared; then switch to the buffer with the largest b_i (buffer level).

   Clear the Largest Work (CLW): Do not change setup until the buffer of the parts being processed is cleared; then switch to the buffer with the largest $\tau_i b_i$ (work load).

   Clear a Fraction (CAF): This is a class of policies which generalizes CLB and CLW. Do not change setup until the buffer of the parts being processed is cleared; then switch to the buffer i for which $b_i \geq \varepsilon \sum_j b_j$, For some ε.

   Perkins and Kumar and Kumar and Seidman showed that under this class of policies the system remains stable. That is, the system can operate with finite buffer levels.

   Lou et al. (1991) used the same policies to study a deterministic single-machine-multiple part system. Their main result is a characterization of the dynamics induced by CLW policies and tighter bounds for the buffers (compared to the ones provided by Perkins and Kumar, 1989). Lou et al. (1992) extended their results to the case when the machine is unreliable and showed that the total Work-In-Process (WIP) is a recurrent stochastic process.

1.3.4 Other policies and results

   Connolly (1992) proposed a heuristic for the two-part one-machine setup system, based on known results from the non-setup system. Her approach is based on a local optimization that maximizes the progress toward a target cyclic schedule. Her model considers backlog only.

Bai and Burhanpurwala (1993) studied a stochastic two-part, two-machine, two-buffer manufacturing system with dynamic setup changes. They developed a four-level hierarchical real-time production control algorithm.

Gallego (1993) studied the ELSP problem in the case of a machine subject to disruptions of small magnitude. He shows that the optimal policy selects the production lot sizes as a linear function of the current inventory levels.

Hu and Caramanis (1995) studied a three-part setup system. First, they solved the problem numerically to deduce the structural properties of the optimal policies. Based on the numerical results, they proposed near-optimal policies. The latter are based on primary and secondary setup change surfaces (in n-dimensional backlog/surplus space). When encountered, primary switching surfaces (which are of dimension n-1) indicate that the current setup must be changed. Secondary switching surfaces (which are of lower dimension) indicate which setup to change to. Hu and Caramanis also proposed a procedure for designing primary and secondary switching surfaces.

1.4 Assumptions

Throughout this proposal, we assume that the following assumptions hold:

- The system has infinite supply of raw material
- The system has infinite space available for the storage of finished parts;
- Many copies of each product or part are made;
- The parts are grouped into lots or batches of specified sizes;
- The products or parts are not required to be made in a single batch;
- We can decide when exactly to perform a setup change;

- The system produces perfect parts or products (no rework);
- All non-fulfilled orders or tardy jobs are backlogged.

1.5 Dissertation Outline

In Chapters 2, 3, 4, and 5, we study a manufacturing system consisting of a single perfectly reliable machine producing two part types. In Chapter 2, we give the general mathematical formulation of the problem and its solution approach. In Chapter 3, we study the system at the steady state. The optimal steady state solution of the two-part-type system is derived analytically. The formulation is then extended to the multi-part-type system and the optimal solution is obtained numerically by means of a modified classical nonlinear programming algorithm. In Chapter 4, we study the transient behavior of the two-part-type system with negligible setup costs. In Chapter 5, we extend the results of Chapter 4 to study the transient behavior of the more general two-part-system including setup times and setup costs. In Chapter 6, we study a system consisting of a single machine subject to random failures and repairs. The setup times are random and nonresumable. That is, after each repair completion, the machine must be set up for a part type. A stochastic optimal control formulation of the problem is given. The problem is then solved numerically. Based on the numerical solution and results from the deterministic case, a heuristic, implementable in real-time, is proposed for the setup switching policy. Simulation results for testing the heuristic performance are provided. In Chapter 7, we summarize our results and indicate possible future directions of this research work.

MATHEMATICAL FORMULATION AND SOLUTION APPROACH
DETERMINISTIC SYSTEM

2.1 Introduction

The goal of the Production and Setup Scheduling Problem (PSSP) is to determine the optimal production rates and setup epochs of several products on a single machine. The latter is assumed to have controllable production rates (i.e., the production rate can be any value between zero and a certain maximum value). Each product has known constant demand rate and processing time. When production is switched from one product to the next a constant setup time as well as a fixed setup cost are incurred. Backlog is allowed and the system is not necessarily in steady state. The objective is to control the production rate of each product as well as to control the setup change epochs so as to minimize the total setup, inventory and backlog costs over a finite or infinite planning horizon. The problem reduces to the famous Economic Lot Scheduling Problem (ELSP), when the planning horizon is infinite, the system is in steadystate, the machine has fixed production rates, no backlog is allowed, and the objective is to determine lot sizes that minimize the average setup and inventory holding costs per unit time.

The PSSP is concerned with the study of the system during its transient period (i.e., before stabilizing at the steady state) as well as during its steady state period, throughout the planning horizon. For the deterministic system, we consider a finite planning horizon and assume that the steady state is finite time achievable. This model may have many applications in practice. For

instance, consider the following situation. We would like to plan the production of several products, periodically (say quarterly), on a bottleneck machine. The demand rate for each period of time is forecasted at the beginning of that period and should be satisfied during the planning period. The machine has enough capacity to meet the forecasted demand for each product and is fast enough to bring the inventory/backlog of all products to a steady state (which constitutes the most economical way of operating the system). Now, at the end of the period (planning horizon), the inventory and backlog of the products are checked and (possibly) different forecasted demand rates are to be used in the next period. It is clear that the output (inventory/backlog) of the previous period constitutes the input for the current period, thus new initial conditions as well as new parameters (different demand rate) are to be used. As can be seen, at the beginning of each planning period we are faced with the problem of optimally driving the system to its steady state production cycle. Many other factors can change the conditions of the system and might drive it away from its steady state. Thus, the importance of the transient optimal solution.

The PSSP can be formulated as a feedback control problem. The control must respond to certain initial disruptions so as to minimize a certain criterion. This kind of formulation is usually classified under *production flow control* models.

In this chapter, we mathematically formulate and give the solution approach for a deterministic one-machine two-part-type system within a feedback control framework.

2.2 Problem Formulation

Consider a single-machine manufacturing system producing two distinct part types (or products); each has a constant demand rate d_i $(i = 1,2)$. When production is switched from Part Type j to Part Type i $(j \neq i)$, a given constant setup time δ_i and setup cost k_i $(i = 1,2)$ are incurred. Our formulation follows the general framework introduced by Kimemia and Gershwin (1983), where the production flow is modeled as continuous rather than discrete. Let $x_i(t)$ be the production surplus of Part Type i $(i = 1,2)$ at time t. A positive value of $x_i(t)$ represents inventory while a negative value represents backlog. Let $u_i(t)$ be the controlled production rate of the machine producing Type i parts at time t. Let $\sigma(t) = (\sigma_1(t), \sigma_2(t), \sigma_{12}(t), \sigma_{21}(t))$ be the setup state vector of the machine at time t. Where, $\sigma_i(t)$, $\sigma_{ij}(t)$ $(j \neq i, i = 1,2, j = 1,2)$ are right continuous binary functions of t, such that $\sigma_i(t) = 1$ when the machine is ready to produce Type i parts and $\sigma_i(t) = 0$ otherwise. $\sigma_{ij}(t) = 1$ when the machine is undergoing a setup change from Part Type j to Part Type i and $\sigma_{ij}(t) = 0$ otherwise. Let $s(t)$ be a nonnegative right continuous function of t which takes on the value δ_i at the beginning of each setup change to Part Type i $(i = 1,2)$ and decreases with time. $s(t)$ indicates whether a setup is completed or not. We assume that initially the machine is not set up for either part type.

System dynamics and constraints

The dynamics of the system can be described by

$$\frac{dx_i(t)}{dt} = u_i(t) - d_i, \ i = 1,2; \tag{2.1}$$

$$0 \leq u_i(t) \leq U_i \sigma_i(t), \ i = 1,2 \tag{2.2}$$

where U_i is the maximum production rate of the machine when it is producing Type i parts.

The setup states of the machine obey the following set of constraints:

$$\sigma_1(t) + \sigma_2(t) + \sigma_{1,2}(t) + \sigma_{2,1}(t) = 1; \tag{2.3}$$

$$\text{if } \sigma_i(t^-) = 1 \text{ and } \sigma_i(t) = 0, \text{ then } s(t) = \delta_j \text{ and } \sigma_{ij}(t) = 1; \tag{2.4}$$

$$\text{if } s(t^-) > 0 \text{ and } \sigma_{ij}(t^-) = 1, \text{ then } \dot{s}(t) = -1 \text{ and } \sigma_{ij}(t) = 1; \tag{2.5}$$

$$\text{if } s(t^-) = 0 \text{ and } \sigma_{ij}(t^-) = 1, \text{ then } \sigma_{ij}(t) = 0 \text{ and } \dot{s}(t) = 0 \text{ and } \sigma_j(t) = 1 \tag{2.6}$$

for $i = 1,2$, $j = 1,2$, $i \neq j$, where $\dot{s}(t)$ denotes the time derivative of $s(t)$.

The above equations, (2.3)-(2.6), merely state that if $\sigma_i(t) = 1$, we can either continue producing Part Type i, or decide to switch production to Part Type j. In the latter case, we must spend exactly δ_j amount of time setting up the machine for Part Type j. That is, $\sigma_{ij}(t) = 1$ for exactly δ_j amount of time $(0 \leq s(t) \leq \delta_j)$. After the setup change, $\sigma_j(t) = 1$ and the machine is ready to produce Part Type j.

<u>Penalty function</u>

For mathematical convenience, we assume that setup costs are incurred at a constant rate $\chi_i = k_i / \delta_i$ $(i = 1,2)$ dollars per unit time during a setup change. Hence, at the end of a setup change to Part Type i, we would have a total setup cost of k_i. The instantaneous cost which penalizes the production for being ahead of (i.e., $x_i > 0$) or being behind (i.e., $x_i < 0$) the demand is given by

$$h(x) = \sum_{i=1}^{i=2}\left(c_i^+ x_i^+(t) + c_i^- x_i^-(t)\right)$$

where c_i^+ and c_i^- are the per unit instantaneous inventory holding and backlog costs respectively, and $x_i^+(t) = \max\{x_i(t), 0\}$ and $x_i^-(t) = \max\{-x_i(t), 0\}$. The total instantaneous cost is then given by

$$g(x, \sigma) = h(x) + \sum_{i=1, j\neq i}^{i=2} \chi_i \sigma_{ji}(t) = \sum_{i=1, j\neq i}^{i=2}\left(c_i^+ x_i^+(t) + c_i^- x_i^-(t) + (k_i/\delta_i)\sigma_{ji}(t)\right). \tag{2.7}$$

State variable

The state variable of the system is given by $x(t) = (x_1(t), x_2(t))$. The state space is given by $x \in \Re^2$.

Control variables

The control vector $u(t) = (u_1(t), u_2(t))$ is the vector of production rates of the two part parts. The control vector $\sigma(t) = (\sigma_1(t), \sigma_2(t), \sigma_{12}(t), \sigma_{21}(t))$ is the vector of setup states of the system. We denote by (σ, u) the complete control vector.

Production control space

The production control space represents the set of all feasible production rates. When the setup state is $\sigma(t)$ at time t, it is given by

$$\Omega(\sigma(t)) = \{u(t) \mid 0 \le u_i(t) \le U_i . \sigma_i(t), \ i = 1,2\}. \tag{2.8}$$

Hence, for each setup state we have a different production control space. These are given by

$$\Omega(0,0) = \{u(t) \mid u_i(t) = 0, \ i = 1,2\};$$
$$\Omega(1,0) = \{u(t) \mid 0 \le u_1(t) \le U_1, \ u_2(t) = 0\}; \tag{2.9}$$
$$\Omega(0,1) = \{u(t) \mid u_1(t) = 0, \ 0 \le u_2(t) \le U_2\}.$$

Usually the production control space is referred to as the capacity set.

Setup control space

The setup control space is the set of all possible setup vectors $\sigma(t) = (\sigma_1(t), \sigma_2(t), \sigma_{12}(t), \sigma_{21}(t))$ satisfying constraints (2.3)–(2.6). Let Φ denote this set.

Admissible control policies

We denote by $\Xi(\Phi, \Omega)$ the set of feasible controls. The set of admissible control policies A is the set of all mappings μ, $\mu : \Re^2 \rightarrow \Xi(\Omega, \Phi)$, which satisfy $\mu(x) = (\sigma, u)$ and which are piecewise continuously differentiable. These

admissible control policies are feedback controls that specify the control actions (setup and production rate of the machine) to be taken, given the state of the system.

Objective function

The objective is to determine an optimal control policy $\mu^* \in A$, corresponding to a setup control $\sigma^* = (\sigma_1^*, \sigma_2^*, \sigma_{12}^*, \sigma_{21}^*)$ and a production flow rate control $u^* = (u_1^*, u_2^*)$, that minimizes for each initial state $x(t)$ the following cost functional:

$$J_\mu(x(t), t) = \int_t^{t_f} g(x(s), \sigma(s)) ds \tag{2.10}$$

where, the minimization is over all functions $\mu(x(\tau)) = (\sigma(\tau), u(\tau))$ such that $x(\tau)$, $\sigma(\tau)$, and $u(\tau)$ satisfy the system dynamics (2.1) and $(\sigma(\tau), u(\tau)) \in \Xi(\Phi, \Omega)$ for $t \leq \tau \leq t_f$, where t_f is assumed to be sufficiently large.

2.3 Solution Approach

The optimal solution to the above problem can be obtained in two parts by dividing the planning horizon ($[t, t_f]$) into two distinct periods. A transient period ($[t, t_s]$) and a steady state period ($[t_s, t_f]$). The steady state period corresponds to the case where the state of the system (inventory/backlog or production surplus levels) has already reached a cyclic schedule, where the produced lots for each part type are of constant size over time. The transient period corresponds to the period where the state of the system still has not reached the cyclic schedule. Formal definitions of the steady state and transient solutions for our problem are given as follows:

Definition 2.1: When the production surplus trajectory follows a cyclic schedule in x-space (or production surplus space), we say that the system has reached a steady state.

Definition 2.2: An optimal steady state solution corresponds to a cyclic trajectory in x-space minimizing the total average setup, inventory and backlog costs per time unit.

Definition 2.3: A transient solution is defined as a trajectory in x-space emanating from an initial point and reaching the steady state in a finite amount of time.

Definition 2.4: An optimal transient solution is a transient solution that minimizes the cost of reaching the steady state. That is, among all transient trajectories that lead to the cyclic schedule, it is the one that gives the minimum total cost (setup, inventory and backlog costs).

In this section, we state and prove a theorem that will allow us to reduce the set of feasible production rates from an infinite set to a finite one with only three possible production rates.

Since we distinguish between the transient and the steady state periods, the cost functional given by (2.10) can be written as the sum of two components: One is due to the transient period, and the other is due to the steady state period. Let t_s be the time instant the system reaches the steady state. The total cost can then be written as follows:

$$J_\mu(x(t),t) = \int_t^{t_s} g(x(s),\sigma(s))ds \; + \; \int_{t_s}^{t_f} g(x(s),\sigma(s))ds \,;$$
$$= J_\mu^T(x(t),t) \; + \; (t_f - t_s)J_\mu^S(x(t_s),t_s). \tag{2.11}$$

We refer to $J_\mu^T(x(t),t)$ as the transient cost,

$$J_\mu^T(x(t),t) = \int_t^{t_f} g(x(s),\sigma(s))ds; \qquad (2.12)$$

And $J_\mu^S(x(t_s),t_s)$ as the average steady state cost,

$$J_\mu^S(x(t_s),t_s) = \frac{1}{(t_f - t_s)}\int_{t_s}^{t_f} g(x(s),\sigma(s))ds. \qquad (2.13)$$

Throughout this dissertation, we assume that the following condition holds.

<u>A1:</u> We assume that $(t_f - t)$, the planning horizon, is long enough so that the system reaches the steady state and stays there for a long period of time.

The following Lemma is based on Assumption A1:

<u>Lemma 2.1:</u> Let t_s be the time the system reaches the steady state, and Let T be the length of the cyclic schedule. Then, we have

$$\lim_{(t_f - t_s) \to \infty} \frac{1}{t_f - t_s}\int_{t_s}^{t_f} g(x(s),\sigma(s))ds = \frac{1}{T}\int_0^T g(x(s),\sigma(s))ds;$$

And by assumption A1, we have

$$J_\mu^S(x(t_s),t_s) \cong \frac{1}{T}\int_0^T g(x(s),\sigma(s))ds.$$

<u>*Proof:*</u> Let $t_f - t_s = nT + t_0$, where n is a nonnegative integer. Then,

$$\frac{1}{t_f - t_s}\int_{t_s}^{t_f} g(x(s),\sigma(s))ds = \frac{1}{nT + t_0}\int_{t_s}^{t_s + nT + t_0} g(x(s),\sigma(s))ds;$$

$$= \frac{n}{nT + t_0}\int_0^T g(x(s),\sigma(s))ds + \frac{1}{nT + t_0}\int_{t_s}^{t_s + t_0} g(x(s),\sigma(s))ds$$

giving

$$\lim_{(t_f - t_s) \to \infty} \frac{1}{t_f - t_s}\int_{t_s}^{t_f} g(x(s),\sigma(s))ds = \lim_{n \to \infty}\left\{\frac{n}{nT + t_0}\int_0^T g(x(s),\sigma(s))ds + \frac{1}{nT + t_0}\int_{t_s}^{t_s + t_0} g(x(s),\sigma(s))ds\right\}.$$

Since $\int_0^T g(x(s), \sigma(s))ds$ and $\int_{t_s}^{t_s+t_0} g(x(s), \sigma(s))ds$ are finite it follows that

$$\lim_{(t_f - t_s) \to \infty} \frac{1}{t_f - t_s} \int_{t_s}^{t_f} g(x(s), \sigma(s))ds = \frac{1}{T} \int_0^T g(x(s), \sigma(s))ds.$$

Now by assumption A1, we have

$$J_\mu^S(x(t_s), t_s) \cong \lim_{(t_f - t_s) \to \infty} \frac{1}{t_f - t_s} \int_{t_s}^{t_f} g(x(s), \sigma(s))ds = \frac{1}{T} \int_0^T g(x(s), \sigma(s))ds. \blacksquare$$

The following theorem will be very important for the solution of the deterministic system.

<u>Theorem 2.1:</u> The optimal production rate vector $u^*(s) = (u_1^*(s), u_2^*(s))$, $(t \leq s \leq t_f)$, belongs to the finite set of vectors $\Omega^* = \{(0,0), (U_1, 0), (d_1, 0), (0, U_2), (0, d_2)\}$, where $u_i^*(s) = 0$ if the machine is not producing or undergoing a setup change. The machine produces at the demand rate $(u_i^*(s) = d_i)$ only when $x_i(s) = 0$ $(i = 1, 2)$.

<u>*Proof:*</u> First, we prove the theorem for the transient case. Then, we show that a similar argument can be used for the steady state case. The proof is based on the Hamilton–Jacobi–Bellman (HJB) equation. Throughout the proof, we assume that the optimal cost functional is differentiable in x and t. In fact, we will show later in the dissertation that the optimal state trajectory is continuous and piecewise linear. Hence, the optimal cost will not depend explicitly on t and will be the sum of quadratics in x (since the cost rate is linear in x) and therefore differentiable in x.

Let $J_{T_f}^*(x, t) = \min_{\sigma, u \in (\Phi, \Omega)} \int_t^{T_f} g(x(s), \sigma(s))ds$, where T_f is finite.

<u>Transient case:</u> In this case, we let $T_f = t_e$ in the expression of $J_{T_f}^*$ to obtain the optimal transient cost component. The HJB equation (see Gershwin (1993) for a formal derivation) is given by:

$$-\frac{\partial J_{T_f}^*(x,t)}{\partial t} = \min_{\sigma, u \in \{\Phi, \Omega\}} \left\{ g(x,\sigma) + \frac{\partial J_{T_f}^*(x,t)}{\partial x_1}(u_1 - d_1) + \frac{\partial J_{T_f}^*(x,t)}{\partial x_2}(u_2 - d_2) \right\}.$$

It is clear that, when the machine is undergoing a setup change to a Part Type, there is no decision to make and (u_1^*, u_2^*) is forced to be equal to $(0,0)$. Now, assume that we know the optimal setup state of the machine. Let $\sigma = (1,0,0,0)$ be this setup state. That is, the machine can produce Part Type 1. In this case, the HJB equation can be rewritten as follows:

$$-\frac{\partial J_{T_f}^*(x,t)}{\partial t} = \min_{u \in \Omega(1,0)} \left\{ g(x,\sigma) + \frac{\partial J_{T_f}^*(x,t)}{\partial x_1}(u_1 - d_1) + \frac{\partial J_{T_f}^*(x,t)}{\partial x_2}(u_2 - d_2) \right\}.$$

Notice that at each time instant t, if we knew $J_{T_f}^*(x,t)$, we would solve a linear programming problem for which u_1 and u_2 are the decision variables, $\partial J_{T_f}^*/\partial x_1$ and $\partial J_{T_f}^*/\partial x_2$ are the cost coefficients and $\Omega(1,0)$ is the constraints set. $\Omega(1,0) = \{(u_1,u_2) | 0 \le u_1 \le U_1, u_2 = 0\}$ is bounded and convex. We know that the solution of the above linear programming problem is always at a vertex of the constraint set $\Omega(1,0)$. That is, (u_1^*, u_2^*) is either equal to $(0,0)$ (if $\partial J_{T_f}^*/\partial x_1 > 0$) or equal to $(U_1,0)$ (if $\partial J_{T_f}^*/\partial x_1 < 0$). Furthermore, the solution is unique if the cost coefficient $\partial J_{T_f}^*/\partial x_1$ is nonzero. In the case $\partial J_{T_f}^*/\partial x_1 = 0$, the solution is not unique anymore since any (u_1^*, u_2^*) will not affect the objective function of the linear programming problem at time instant t. However, to keep the cost coefficient $\partial J_{T_f}^*/\partial x_1$ equal to zero at time instant $t + \delta t$, we should produce Part Type 1 at the demand rate d_1 so as to minimize the rate of increase of the cost

function $J^*_{T_f}$. In this case (u^*_1, u^*_2) is equal to $(d_1, 0)$. A similar argument is used when the optimal setup state is $\sigma = (0,1,0,0)$.

__Steady-state case:__ In this case, let $t_s = t$ and $T_f = t_f$ in the expression of $J^*_{T_f}$ above. Let

$$J^* = \min_\mu J^s_\mu(x,t) = \min_{\sigma, u \in (\Phi, \Omega)} \frac{1}{(t_f - t_s)} \int_{t_s}^{t_f} g(x(s), \sigma(s)) ds.$$

Here, we have an average cost formulation. The HJB equation (see Kushner and Dupuis (1992) for a formal derivation) in this case is given by

$$J^* = \min_{\sigma, u \in (\Phi, \Omega)} \left\{ g(x, \sigma) + \frac{\partial V(x,t)}{\partial x_1}(u_1 - d_1) + \frac{\partial V(x,t)}{\partial x_2}(u_2 - d_2) \right\}$$

where

$$V(x,t) = \lim_{(t_f - t) \to +\infty} J^*_{T_f} - (t_f - t) J^*.$$

As in the previous case, for each time instant t, if we knew $V(x,t)$, we would solve a linear programming problem which decision variables are the production rates u_1 and u_2 and which cost coefficients are $\partial V/\partial x_1$ and $\partial V/\partial x_2$ respectively. Hence, using a similar argument as for the transient case and using Lemma 1, the result follows immediately. ∎

The general approach for the deterministic system consists of determining the steady state optimal solution first (which is in principal easier to get than the transient one), which corresponds to an optimal cyclic schedule in x-space. Based on Lemma 2.1, the steady state optimal solution consists of solving a special economic lot scheduling problem (ELSP) where backlog is allowed and the production rates are controllable. To obtain the transient optimal solution, we partition the x-space into mutually exclusive regions. For initial production surplus vectors $x(t)$ in each region, we determine an optimal

trajectory emanating from $x(t)$ and reaching the cyclic schedule (which becomes a target set in this case) with minimum total cost and in finite amount of time. Each region will be characterized by a family of optimal trajectories leading to the cyclic schedule.

CHAPTER 3
OPTIMAL STEADY STATE SOLUTION
DETERMINISTIC SYSTEM

3.1 Introduction

As mentioned in Chapter 2, the steady problem corresponds to a special Economic Lot Scheduling Problem (ELSP) where, in addition, the production rates are controllable and backlog (negative production surplus) is allowed. In the following, we give a brief background of the ELSP due to its importance as a research problem.

The ELSP deals with the scheduling of the production of several products on one or more identical machines. Each product has a known constant demand rate and a fixed production rate when it is being produced. When production is switched from one product to the next a sequence-independent constant setup time as well as a fixed setup cost are incurred. The time horizon is infinite, the system is in steady state and no backlog is allowed. The objective is to determine lot sizes that minimize the average setup and inventory holding costs per unit time. In Gallego (1989), the ELSP is extended to allow backlog. In this chapter, we consider the case where nonsatisfied demand is totally backlogged.

Although there has been a large amount of research work on the ELSP, an optimal solution approach has not been proposed yet. Rather, good (some times excellent) heuristics have been suggested. A comprehensive review of the ELSP through 1976 is given in Elmaghraby (1978). Recent work on the ELSP includes the work of Boctor (1982), Hsu (1983), Maxwell and Singh (1983),

Axsäter (1984), Goyal (1984), Roundy (1985), Dobson (1987), Gallego (1989), Jones and Inman (1989), Lee and Surya (1989), Carreno (1990), and Zipkin (1991). Most of the aforementioned work emphasized the feasibility of cyclic schedules. A cyclic schedule can be one of two types: Common Cycle schedule or Basic Period schedule. The Common Cycle schedule is also known as the Rotation Cycle schedule, where production is cycled through the products every T units of time. It is well known that this type of schedule is always feasible. Basically, for the Common Cycle schedule, the sequencing problem is eliminated. In the Basic Period schedule, each product's cycle time is an integer multiple of a basic cycle time. Since the production rates are constant, all lots of each product are of equal size. Usually, the Basic Period schedule gives a lower cost than the Common Cycle schedule. The main problem of the Basic Period schedule is its feasibility. In fact, for this kind of approach, Hsu (1983) has shown that even the problem of finding a feasible schedule is an NP-hard problem. To overcome this feasibility problem, Dobson (1987) suggested a new formulation that allows time varying production runs and which includes setup times explicitly in the problem formulation. The main advantage of this approach is that it always provides a feasible schedule.

Recently, researchers have dropped the assumption of fixed production rates at the machine and have improved the ELSP model using controllable production rates. The production rate of each product can be chosen to be any value less than a maximum rate. Recent work along this new direction includes the work of Buzacott and Ozkarahan (1983), where they studied the case of a two-product system. First, they categorize the products according to their dollar value of usage, then they showed that only the product with the lower dollar value of usage is produced at maximum. Arizono et al. (1989) studied the effects of controllable production rates on inventory systems. They showed

that a controllable production rate inventory system is more efficient than one with fixed production rates. Silver (1990) studied a multi-product system under the Common Cycle schedule assumption. He showed that at most one product slows down its production rate. The optimal production rates and cycle time length were obtained numerically. In all of the aforementioned work, the production rates were decided at the beginning and supposed to be fixed during a product's production run. Moon et al. (1991) generalized Silver's model by considering a system with controllable production rates during the production runs of the products. They obtained the optimal production rates numerically and showed that savings almost twice as large as those reported in the literature can be obtained.

The chapter is organized as follows. In Section 3.2, we present the notation and state the assumptions of the model. In Section 3.3, we study the two-product problem. In Section 3.4, we extend the formulation to the multi-product problem and present an algorithm to obtain the optimal solution numerically. We conclude the chapter with Section 3.5.

3.2 The Two-Part-Type Case

In Chapter 2, it has been shown that the steady state period corresponds to the case where the state of the system (production surplus levels) has already reached a cyclic schedule, where the produced lots for each part type are of constant size over time.

If the machine were perfectly flexible (i.e., with zero setup change times and costs), it would be optimal to produce both parts simultaneously at the demand rates (theoretically, it is optimal to instantaneously switch production back and forth between the two products, since no cost and time are incurred

when we switch). Thus, keeping the production surplus at the zero level. In the case of significant setup times and/or costs, it is not possible to produce both part types at the same time. Thus, at the steady state the production surplus vector must follow a cyclic schedule that repeats itself over time.

Based on Theorem 2.1 defined in Chapter 2, it is not difficult to see that a feasible cyclic schedule has the general structure shown in Figure 3.1, where x_i represents the inventory/backlog axis of Part Type i ($i = 1,2$). Now, assume that we start the cycle at Point $P1$, we then progress toward Point $P2$ by producing Part Type 1 at the demand rate along segment $[P1,P2]$, where x_2 decreases until we reach Point $P2$. At this point, we increase the production rate to the maximum and continue producing Part Type 1. x_1 increases while x_2 decreases until we reach Point $P3$, where we switch production to Part Type 2. During the setup of the machine for Part Type 2, both inventory levels decrease. Once the machine is ready to produce Part Type 2 (Point $P4$), the production begins with the maximum allowable rate in order to eliminate backlog as soon as possible (since after the setup, we endup with a backlog for Part Type 2). Once Part Type 2 backlog is completely eliminated, we decrease the production rate to the demand rate so that the system moves along $[P5,P6]$. When we reach $P6$, the production rate of Part Type 2 is increased to the maximum and a certain inventory is built to hedge against future shortages brought about setups and production of Part Type 1. At Point $P7$, we setup the machine for Part Type 1. At Point $P8$, we produce Part Type 1 at maximum production rate to eliminate backlog as soon as possible until we reach Point $P1$, where we start the cycle all over again.

The steady state solution is completely characterized when the optimal location and shape (in x-space) of the cyclic schedule is known. Using Lemma 2.1 of Chapter 2, it can be seen that determining the cyclic schedule is

equivalent to solving a special two-product ELSP. Graphically, the optimization problem can be seen as one of locating and determining the shape of the cyclic schedule in the x-space so as to minimize the average setup, inventory holding and backlog costs per unit time. In the following, we formulate the problem mathematically and then derive the optimal solution in closed form. Notice that we do not consider idle times. The reason is as follows: the purpose of idling the machine (i.e., stopping the machine completely) during the production of a product is usually to stretch the cycle time and delay setup costs which may be high in some cases. In our case, this task is accomplished by producing the products at the demand rate which eliminates inventory and backlog costs for a product during its production at the demand rate along with delaying setup costs. In the case of fixed production rates, during idle times there is a certain inventory or backlog that is present for which a cost is incurred. Therefore idling cannot be optimal in our case.

3.2.1 Problem formulation

Notation:

For Product i $(i = 1,2)$

t_i: time spent producing at maximum rate

τ_i: time spent producing at the demand rate

S_i: maximum inventory level

s_i: maximum backlog level

$\gamma_i = c_i^+ c_i^- / (c_i^+ + c_i^-)$: cost factor

$\rho_i = d_i / U_i$: utilization factor of the machine by Product i

$T = \sum_{i=1}^{i=2} (t_i + \tau_i + \delta_i)$: duration of the cyclic schedule

$\delta = \delta_1 + \delta_2$: total setup time during T

Figure 3.1: General Structure of the Cyclic Schedule

$K = k_1 + k_2$: total setup cost during T

$\rho = \rho_1 + \rho_2$: total utilization factor of the machine

$\alpha_i = (1 - \rho_i)/(1 - \rho)$

$A_i = \gamma_i/2d_i(1 - \rho_i)$

Without loss of generality, let us start the cyclic schedule shown in Figure 3.1 at either Points $P1$ or $P5$, and let us denote by i the index of the product we start with.

According to Figure 3.1, the inventory of Product i behaves as shown in Figure 3.2. S_i and s_i are the maximum inventory and backlog levels respectively attained within a cycle of Product i. Note that s_i is negative and S_i is nonnegative. Let $Q_i = S_i - s_i$. This quantity is known as the replenishment quantity in the inventory theory literature. Obviously, Q_i must be positive.

Notice here that the production cycle is completely characterized when the quantities S_i, Q_i and τ_i are determined. This, we do in the following.

Based on Figure 3.2, it is not difficult to see that the total inventory holding cost and the total backlog cost over a cycle are given by

$$Inventory\ cost = \frac{1}{2} c_i^+ \frac{S_i^2}{d_i(1-\rho_i)};$$

$$Backlog\ cost = \frac{1}{2} c_i^- \frac{s_i^2}{d_i(1-\rho_i)} = \frac{1}{2} c_i^- \frac{(S_i-Q_i)^2}{d_i(1-\rho_i)}.$$

The average setup, inventory holding and backlog cost per unit time of Product i is then given by

$$F_i(S_i,Q_i) = \frac{k_i}{T} + \frac{1}{2}\frac{c_i^+}{T}\frac{S_i^2}{d_i(1-\rho_i)} + \frac{1}{2}\frac{c_i^-}{T}\frac{(S_i-Q_i)^2}{d_i(1-\rho_i)}.$$

Figure 3.2: Inventory behavior of Part Type i.

The total average setup, inventory holding and backlog cost per unit time is given by

$$F(S_1, S_2, Q_1, Q_2) = \sum_{i=1}^{2} \left(\frac{k_i}{T} + \frac{1}{2} \frac{1}{d_i(1-\rho_i)} \left(\frac{c_i^+ S_i^2 + c_i^- (S_i - Q_i)^2}{T} \right) \right).$$
(3.1)

In the following, we show how T and Q_i are related to τ_i, the time spent producing Product i at the demand rate. First, it is not difficult to see that

$$t_i = \frac{Q_i}{d_i} \frac{\rho_i}{(1-\rho_i)}.$$
(3.2)

From the demand satisfaction constraint $Td_i = \tau_i d_i + t_i p_i$, we get

$$Q_i = d_i(1-\rho_i)(T - \tau_i).$$
(3.3)

But $T = \sum_{i=1}^{i=2}(t_i + \tau_i + \delta_i)$. Substituting for t_i and rearranging terms, we get

$$T = \frac{1}{(m-1)} \sum_{i=1}^{i=m} \frac{Q_i}{d_i} - \frac{\delta}{(m-1)} = \sum_{i=1}^{m} \frac{(1-\rho_i)}{(1-\rho)} \tau_i + \frac{\delta}{(1-\rho)}.$$
(3.4)

Here, $m = 2$ (two products). Substituting T in the expression of Q_i, gives

$$Q_i = q_i \left(1 + \sum_{j=1}^{m}(1-\rho_j)\frac{\tau_j}{\delta} - (1-\rho)\frac{\tau_i}{\delta} \right)$$
(3.5)

Where $q_i = \frac{d_i(1-\rho_i)}{(1-\rho)}\delta$. q_i can be seen as the replenishment quantity when the τ_i's are all zero. It is clear that the independent variables of the model are the S_i's and the τ_i's. Letting S and τ be the vectors which components are the S_i's and the τ_i's respectively, the total average cost per unit time can be rewritten as follows:

$$F(S, \tau) = \frac{\sum_{i=1}^{i=m} \left(k_i + \frac{1}{2} \frac{1}{d_i (1 - \rho_i)} \left(c_i^+ S_i^2 + c_i^- (S_i - Q_i)^2 \right) \right)}{\frac{1}{(m-1)} \sum_{i=1}^{i=m} \frac{Q_i}{d_i} - \frac{\delta}{(m-1)}}.$$ (3.6)

The minimization problem can be stated as follows:

(P)   *Minimize* $F(S, \tau)$

   Subject to

$$Q_i = q_i \left(1 + \sum_{j=1}^{m} (1 - \rho_j) \tau_j / \delta - (1 - \rho_i) \tau_i / \delta \right), \quad i = 1, \ldots, m;$$

$$S_i \geq 0, \ \tau_i \geq 0, \ Q_i \geq 0, \ i = 1, \ldots, m.$$

3.2.2 Optimal solution

Our solution approach consists of two steps. First, we obtain the optimal S_i's as a function of the τ_i's through the quantities Q_i. Second, we ignore the nonnegativity constraints and solve for the τ_i's implicitly through the quantities Q_i. If the obtained solution is feasible (i. e., satisfies the nonnegativity constraints), then it is optimal. If the nonnegativity constraints are violated, then in the optimal solution of the constrained problem, either one or both nonnegativity constraints will be effective. Hence, we must distinguish the three possible constrained solutions. That is, $(\tau_1 = 0, \tau_2 > 0)$, $(\tau_1 > 0, \tau_2 = 0)$, or $(\tau_1 = 0, \tau_2 = 0)$. A detailed procedure for obtaining the feasible optimal solution will be provided later in the section.

Setting the partial derivatives of $F(S, \tau)$ with respect to S_i to zero and rearranging terms gives

$$S_i^* = \frac{c_i^-}{c_i^+ + c_i^-} Q_i, \quad i = 1, \ldots, m.$$ (3.7)

The second partial derivative of $F(S, \tau)$ with respect to S_i is always positive. Therefore, $F(S, \tau)$ is strictly convex in S_i and S_i^* given above is a global minimum for $F(S, \tau)$. Substituting S_i^* in $F(S, \tau)$ gives

$$F(\tau) = \frac{K + A_1 Q_1^2 + A_2 Q_2^2}{Q_1/d_1 + Q_2/d_2 - \delta} \tag{3.8}$$

Before we proceed with the solution for the τ_i's, let us prove the following proposition:

<u>Proposition 3.1</u> $F(\tau)$ is strictly convex in τ_1 and τ_2.

<u>Proof:</u> The Hessian matrix of $F(\tau)$ is given by

$$H = \frac{2}{T^3} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}.$$

Where

$$H_{11} = Ka^2 + A_1 a^2 d_1^2 ((1 - \rho_2)\tau_2 + \delta)^2 + A_2 b^2 d_2^2 (1 - \rho_1)^2 \tau_2^2;$$

$$H_{22} = Kb^2 + A_1 a^2 d_1^2 (1 - \rho_2)^2 \tau_1^2 + A_2 b^2 d_2^2 ((1 - \rho_1)\tau_1 + \delta)^2;$$

$$H_{12} = H_{21} = Kab - A_1 a^2 d_1^2 (1 - \rho_2)\tau_1 ((1 - \rho_1)\tau_2 + \delta) - A_2 b^2 d_2^2 (1 - \rho_1)\tau_2 ((1 - \rho_2)\tau_1 + \delta);$$

$$a = \frac{q_1}{\delta d_1} \text{ and } b = \frac{q_2}{\delta d_2}.$$

It is clear that H_{11} is positive. The determinant of H is given by

$$\det H = \frac{4KA_1 a^2 d_1^2 [a(1 - \rho_1)\tau_1 + b((1 - \rho_2)\tau_2 + \delta)]^2}{T^6}$$

$$+ \frac{4KA_2 b^2 d_2^2 [a((1 - \rho_1)\tau_1 + \delta) + b(1 - \rho_2)\tau_2]^2}{T^6}$$

$$+ \frac{4d_1^2 d_2^2 A_1 A_2 a^2 b^2 [(1 - \rho_1)(1 - \rho_2)\tau_1 \tau_2 - ((1 - \rho_1)\tau_1 + \delta)((1 - \rho_2)\tau_2 + \delta)]^2}{T^6}$$

which is clearly positive. Hence, all the minors of H are positive, and H is positive definite. The result follows immediately. ∎

As a consequence of Proposition 3.1, $F(\tau)$ has a unique global minimum τ^*. In the following, we provide the optimal solution for the cases where $(\tau_1 \geq 0, \tau_2 \geq 0)$, $(\tau_1 = 0, \tau_2 \geq 0)$, $(\tau_1 \geq 0, \tau_2 = 0)$ and the case $(\tau_1 = 0, \tau_2 = 0)$.

Case $(\tau_1 \geq 0, \tau_2 \geq 0)$:

Taking the partial derivatives with respect to τ_1 and τ_2, setting to zero and rearranging terms gives the following system of two nonlinear equations:

$$
\begin{cases}
A_1(2\rho_2 - 1)Q_1^2 + A_2(1 - 2\rho_2)Q_2^2 + 2(\rho_2(d_1/d_2)A_1 + (1 - \rho_2)(d_2/d_1)A_2)Q_1Q_2 \\
\qquad - 2\delta(A_1 d_1 \rho_2 Q_1 + A_2 d_2(1 - \rho_2)Q_2) - K = 0 \\
A_2(2\rho_1 - 1)Q_2^2 + A_1(1 - 2\rho_1)Q_1^2 + 2(\rho_1(d_2/d_1)A_2 + (1 - \rho_1)(d_1/d_2)A_1)Q_1Q_2 \\
\qquad - 2\delta(A_2 d_2 \rho_1 Q_2 + A_1 d_1(1 - \rho_1)Q_1) - K = 0
\end{cases}
\tag{3.9}
$$

<u>Lemma 3.1:</u> The linear equation $A_1 d_1 Q_1 = A_2 d_2 Q_2$ is a solution of the above system.

proof: substituting $Q_1 = (A_2/A_1)(d_2/d_1)Q_2$ in the first equation of (3.9) gives the second equation. ∎

Solving (3.9) using Lemma 3.1, gives the following optimal values of Q_1 and Q_2 for the unconstrained problem (i.e., without the nonnegativity constraints):

$$
Q_1^u = \frac{q_1 d_2 \gamma_2 \left(1 + \sqrt{1 + 2\dfrac{K}{\delta^2}\left(\dfrac{(1-\rho_1)}{\gamma_1 d_1} + \dfrac{(1-\rho_2)}{\gamma_2 d_2}\right)}\right)}{\left(\dfrac{(1-\rho_2)}{(1-\rho)}d_1\gamma_1 + \dfrac{(1-\rho_1)}{(1-\rho)}d_2\gamma_2\right)};
\tag{3.10a}
$$

$$Q_2^u = \frac{q_2 d_1 \gamma_1 \left(1 + \sqrt{1 + 2\dfrac{K}{\delta^2}\left(\dfrac{(1-\rho_1)}{\gamma_1 d_1} + \dfrac{(1-\rho_2)}{\gamma_2 d_2}\right)}\right)}{\left(\dfrac{(1-\rho_2)}{(1-\rho)}d_1\gamma_1 + \dfrac{(1-\rho_1)}{(1-\rho)}d_2\gamma_2\right)}. \tag{3.10b}$$

The superscript u stands for unconstrained. Note that, at this stage, we have no guarantee that τ_1^u and τ_2^u satisfy the nonnegativity constraints.

Case ($\tau_1 = 0$, τ_2 unconstrained):

A similar procedure as in the previous case leads to the following expression for Q_1^u and Q_2^u.

$$Q_1^u = (1-\rho_1)d_1\sqrt{\frac{\left(K + A_2(\delta d_2)^2\right)}{\left(d_1^2(1-\rho_1)^2 A_1 + d_2^2\rho_1^2 A_2\right)}} \tag{3.11a}$$

$$Q_2^u = d_2\delta + \rho_1 d_2\sqrt{\frac{\left(K + A_2(\delta d_2)^2\right)}{\left(d_1^2(1-\rho_1)^2 A_1 + d_2^2\rho_1^2 A_2\right)}} \tag{3.11b}$$

Case (τ_1 unconstrained, $\tau_2 = 0$):

In a similar fashion, we obtain

$$Q_1^u = d_1\delta + \rho_2 d_1\sqrt{\frac{\left(K + A_1(\delta d_1)^2\right)}{\left(d_2^2(1-\rho_2)^2 A_2 + d_1^2\rho_2^2 A_1\right)}} \tag{3.12a}$$

$$Q_2^u = (1-\rho_2)d_2\sqrt{\frac{\left(K + A_1(\delta d_1)^2\right)}{\left(d_2^2(1-\rho_2)^2 A_2 + d_1^2\rho_2^2 A_1\right)}} \tag{3.12b}$$

Case ($\tau_1 = 0, \tau_2 = 0$)

This case is trivial. Indeed, setting τ_1 and τ_2 to zero in (3.5) gives

$Q_1 = q_1$ and $Q_2 = q_2$.

Lemma 3.2: The following procedure gives the optimal solution of the constrained problem.

Procedure 1

Compute Q_1^u and Q_2^u using (3.10a) and (3.10b);

Compute τ_1^u and τ_2^u as follows:

$$\tau_1^u = \frac{Q_2^u}{d_2} - \frac{Q_1^u \rho_1}{(1-\rho_1)d_1} - \delta; \tag{3.13a}$$

$$\tau_2^u = \frac{Q_1^u}{d_1} - \frac{Q_2^u \rho_2}{(1-\rho_2)d_2} - \delta. \tag{3.13b}$$

IF $\tau_1^u \geq 0$ and $\tau_2^u \geq 0$ THEN

$\qquad \tau_1^* = \tau_1^u, \ \tau_2^* = \tau_2^u, \ Q_1^* = Q_1^u \text{ and } Q_2^* = Q_2^u;$

$\qquad$ *STOP*

ELSE

$\qquad$ *Compute Q_1^u and Q_2^u using (3.11a) and (3.11b);*

$\qquad$ *Compute τ_1^u and τ_2^u Using (3.13a) and (3.13b).*

$\qquad$ *IF $\tau_2^u > 0$ THEN*

$\qquad\qquad$ *Calculate $C_2 = F(\tau_1^u, \tau_2^u)$ using (3.8);*

$\qquad$ *ELSE*

$\qquad\qquad C_2 = \infty.$

$\qquad$ *END*

$\qquad$ *Compute Q_1^u and Q_2^u using (3.12a) and (3.12b);*

$\qquad$ *Compute τ_1^u and τ_2^u Using (3.13a) and (3.13b).*

$\qquad$ *IF $\tau_1^u > 0$ THEN*

$\qquad\qquad$ *Calculate $C_1 = F(\tau_1^u, \tau_2^u)$ using (3.8);*

$\qquad$ *ELSE*

$\qquad\qquad C_1 = \infty.$

END

Let $Q_1^u = q_1$ and $Q_2^u = q_2 \Rightarrow \tau_1^u = 0$ and $\tau_2^u = 0$;

Calculate $C_o = F(\tau_1^u, \tau_2^u)$ using (3.8);

$((\tau_1^*, \tau_2^*), (Q_1^*, Q_2^*)) = \arg\min\{C_o, C_1, C_2\}$;

ENDIF

<u>*Proof:*</u> expressing τ_1 and τ_2 as a function of Q_1^u and Q_2^u, gives

$$\tau_1 = \frac{Q_2^u}{d_2} - \frac{Q_1^u \rho_1}{(1-\rho_1)d_1} - \delta;$$

$$\tau_2 = \frac{Q_1^u}{d_1} - \frac{Q_2^u \rho_2}{(1-\rho_2)d_2} - \delta.$$

Now, if τ_1 and τ_2 are both nonnegative, then the solution of the unconstrained optimization problem is feasible and since $F(\tau)$ is convex in τ_1 and τ_2, it follows that the solution is optimal for the constrained problem as well.

If the nonnegativity constraints are violated, then the optimal solution is obtained by comparing the costs of the three possible cases (effective constraints) and picking the minimum. For the cases where only one of the nonnegativity constraints is effective, if we obtain a nonfeasible solution (i.e., the other constraint must be effective too), we set its cost to infinity. This is because when only one nonnegativity constraint is effective, we set the variable for which the constraint is effective to zero and we solve an unconstrained optimization problem with respect to the other variable. Hence, we might obtain a nonfeasible solution. ∎

Once we obtain the optimal r_1^*, r_2^*, Q_1^*, and Q_2^*, we go back and calculate the optimal S_i, s_i, t_i $(i = 1,2)$ and T, which completely characterizes the optimal cyclic schedule. Points PJ $(J=1,2, \ldots, 8)$ of Figure 3.1 are given as follows:

$$P1 = \begin{pmatrix} 0 \\ S_2 - \delta_1 d_2 + s_1 \rho_1 d_2 / d_1 (1 - \rho_1) \end{pmatrix}; \tag{3.14a}$$

$$P2 = \begin{pmatrix} 0 \\ S_2 - \delta_1 d_2 + s_1 d_2 \rho_1 / d_1 (1 - \rho_1) - \tau_1 d_2 \end{pmatrix}; \tag{3.14b}$$

$$P3 = \begin{pmatrix} S_1 \\ s_2 + \delta_2 d_2 \end{pmatrix}; \tag{3.14c}$$

$$P4 = \begin{pmatrix} S_1 - \delta_2 d_1 \\ s_2 \end{pmatrix}; \tag{3.14d}$$

$$P5 = \begin{pmatrix} S_1 - \delta_2 d_1 + s_2 d_1 \rho_2 / d_2 (1 - \rho_2) \\ 0 \end{pmatrix}; \tag{3.14e}$$

$$P6 = \begin{pmatrix} S_1 - \delta_2 d_1 + s_2 d_1 \rho_2 / d_2 (1 - \rho_2) - \tau_2 d_1 \\ 0 \end{pmatrix}; \tag{3.14f}$$

$$P7 = \begin{pmatrix} s_1 + \delta_1 d_1 \\ S_2 \end{pmatrix}; \tag{3.14g}$$

$$P8 = \begin{pmatrix} s_1 \\ S_2 - \delta_1 d_2 \end{pmatrix} \tag{3.14h}$$

Where s_i, S_i, Q_i and τ_i $(i = 1,2)$ are the optimal values calculated using Procedure 1, (3.7) and $s_i = S_i - Q_i$.

3.2.3 Numerical examples

In this sub-section, we compare our results with other models found in the literature. We show that by allowing a certain production time at the demand rate, we achieve significant savings.

The examples used for the comparison are those proposed by Boctor (1982) to illustrate his algorithm for a two-product Basic Period schedule model. These two examples were then used by Lee and Surya (1989) to compare their new algorithm to Boctor's. Also, we use these two examples with the model of controllable production rates (no production at the demand rate) of Buzacott and Ozkarahan (1983). Notice that no backlog is allowed for the two examples.

The case where backlog is not allowed is obtained by letting c_i^- (the backlog cost rate) go to infinity. Which is basically equivalent to replacing γ_i by c_i^+ for $i = 1,2$ in the above formulas. In this case, the s_i's become all zero.

Example 1

i	d_i	U_i	δ_i	k_i	c_i^+
1	20000	160000	0.0125	15	0.005
2	27000	162000	0.0250	25	0.004

Example 2

i	d_i	U_i	δ_i	k_i	c_i^+
1	3500	100000 .	0.5	2500	0.150
2	46500	100000	0.3	18500	0.005

The optimal solution for the examples is as follows:

Example 1: $\tau_1^* = 0.4282$, $\tau_2^* = 0.4815$, $t_1^* = 0.0900$, $t_2^* = 0.1111$, $T^* = 1.1483$.

Example 2: $\tau_1^* = 6.0248$, $\tau_2^* = 0.0000$, $t_1^* = 0.2521$, $t_2^* = 6.1509$, $T^* = 13.2278$.

Table 3.1 shows the computational results of this comparison. The upper part of each entry in Table 3.1 shows the average cost for each schedule, while the lower part shows how much savings are made with our schedule. For instance, in Example 2, our schedule average cost is 51.3% lower than that of Boctor's. Although we expected the Basic Period schedule to perform better than the Common Cycle schedule, the results show that by introducing a production at the demand rate, we achieve savings up to 66% on the average cost of the schedule. Notice that, even with a controllable production rate but without a production rate at the demand rate (Buzacott and Ozkarahan), we obtain lower average costs.

Table 3.1: Comparison savings for the two-product model

	Our model	Boctor	Lee and Surya	Buzacott and Ozkarahan
Example 1	72.0	118.7	119.2	94.3
	–	64.9%	65.6%	31.0%
Example 2	3403.8	5149.3	4153.0	4144.1
	–	51.3%	22.0%	21.7%

3.2.4 Special cases

As mentioned earlier, based on the optimal solution, we distinguish four different cases. These are shown in Figure 3.3 (a, b, c and d).

In this section, we derive some results that will be useful for Chapters 4 and 5. First, we derive a condition for which we have the case shown in Figure 3.3a. As will be shown, this corresponds to a system where setup costs are negligible and the system is heavily loaded. The transient solution of this case will be studied in Chapter 4. This case constitute the back bone of the optimal transient solution of the more general case studied in Chapter 5. Another

special case is obtained when the setup times are negligible but the setup costs are significant. We will show that in this case τ_1^* and τ_2^* cannot be simultaneously zero.
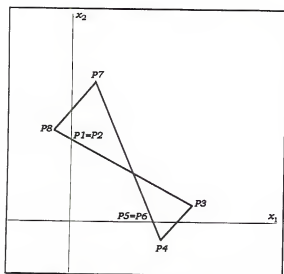

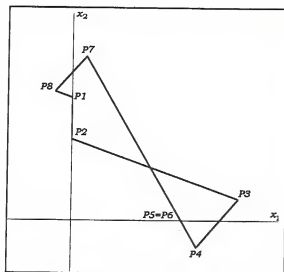Figure 3.3a: $\tau_1 = 0$ and $\tau_2 = 0$.
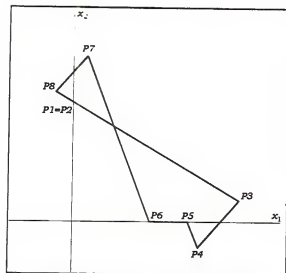
Figure 3.3c: $\tau_1 > 0$ and $\tau_2 = 0$.

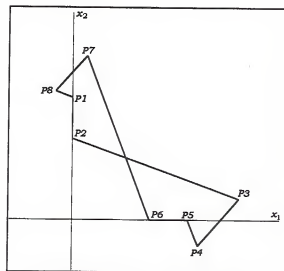Figure 3.3b: $\tau_1 = 0$ and $\tau_2 > 0$.

Figure 3.3d: $\tau_1 > 0$ and $\tau_2 > 0$.

Figure 3.3: Possible shapes of the cyclic schedule

<u>Case of zero setup costs</u>

In the following, we formally derive the condition for which the cyclic schedule in Figure 3.3a is optimal. This condition is given in Theorem 3.1.

<u>Theorem 3.1:</u> Without loss of generality, let Part Type 2 be the part type such that $\gamma_2 d_2 \geq \gamma_1 d_1$. If the setup costs k_1 and k_2 are zero, then $\tau_1^* = 0$ and $\tau_2^* = 0$ if and only if $\rho + \rho_1 - 1 \geq 0$. If $\rho + \rho_1 - 1 < 0$, then $\tau_1^* = 0$ and $\tau_2^* = 0$ if and only if $\gamma_2 d_2 / \gamma_1 d_1 \leq (1 - \rho_1)/(1 - \rho - \rho_1)$.

<u>Proof:</u> The proof is based on the Karush-Kuhn-Tucker (KKT) optimality conditions. Substituting Q_1 and Q_2 in the objective function of Problem (P) above gives the following optimization problem. τ_1 and τ_2 are the decision variables:

$$\textit{Minimize } F(\tau) = \frac{\left(K + \frac{1}{2} H_1 (T_0 + (\alpha_1 - 1)\tau_1 + \alpha_2 \tau_2)^2 + \frac{1}{2} H_2 (T_0 + \alpha_1 \tau_1 + (\alpha_2 - 1)\tau_2)^2 \right)}{T_0 + \alpha_1 \tau_1 + \alpha_2 \tau_2}$$

Subject to

$$\tau_i \geq 0, \ i = 1, 2$$

where

$$H_i = \gamma_i d_i (1 - \rho_i), \ i = 1, 2;$$
$$\alpha_i = (1 - \rho_i)/(1 - \rho), \ i = 1, 2;$$
$$T_0 = \delta/(1 - \rho).$$

Since $F(\tau)$ is strictly convex in τ_1 and τ_2 and the constraint domain is convex, it follows that the KKT optimality conditions are necessary and sufficient (which establishes the if and only if part of the theorem).

Let $g_i(\tau_1, \tau_2) = -\tau_i$ $(i = 1, 2)$. The KKT optimality conditions are given as follows:

$$\nabla F(\dot{\tau_1}, \dot{\tau_2}) + \mu_1 \nabla g_1(\dot{\tau_1}, \dot{\tau_2}) + \mu_2 \nabla g_2(\dot{\tau_1}, \dot{\tau_2}) = 0;$$

$$\mu_1 g_1(\tau_1, \tau_2) = 0;$$

$$\mu_2 g_2(\tau_1, \tau_2) = 0;$$

$$\mu_i \geq 0 \text{ for } i = 1, 2$$

where $\dot{\tau_i}$ is the optimal value of τ_i $(i = 1, 2)$. $\nabla F(\tau_1, \tau_2)$, $\nabla g_1(\tau_1, \tau_2)$ and $\nabla g_2(\tau_1, \tau_2)$ are the gradients of $F(\tau_1, \tau_2)$, $g_1(\tau_1, \tau_2)$ and $g_2(\tau_1, \tau_2)$ respectively.

Letting $\dot{\tau_i} = 0$ $(i = 1, 2)$ and calculating the gradients and substituting in the KKT conditions, gives:

$$\left(\frac{\alpha_1}{2} - 1\right)H_1 + \frac{\alpha_1 H_2}{2} - \frac{\alpha_1 K}{T_0^2} - \mu_1 = 0;$$

$$\frac{\alpha_2 H_1}{2} + \left(\frac{\alpha_2}{2} - 1\right)H_2 - \frac{\alpha_2 K}{T_0^2} - \mu_2 = 0.$$

For the KKT conditions to hold true, we must have $\mu_i \geq 0$ $(i = 1, 2)$. Hence, we must have

$$\frac{\alpha_1 K}{T_0^2} \leq \left(\frac{\alpha_1}{2} - 1\right)H_1 + \frac{\alpha_1 H_2}{2};$$

$$\frac{\alpha_2 K}{T_0^2} \leq \frac{\alpha_2 H_1}{2} + \left(\frac{\alpha_2}{2} - 1\right)H_2.$$

Substituting α_1, α_2, H_1 and H_2 by their expressions gives

$$K \leq \frac{T_0^2 \left((1 - \rho_2)(\gamma_2 d_2 - \gamma_1 d_1) + \rho \gamma_1 d_1\right)}{2};$$

$$K \leq \frac{T_0^2 \left((\rho + \rho_1 - 1)\gamma_2 d_2 + (1 - \rho_1)\gamma_1 d_1\right)}{2}.$$

Letting $K = 0$, we get the following conditions:

$$(1 - \rho_2)(\gamma_2 d_2 - \gamma_1 d_1) + \rho \gamma_1 d_1 \geq 0; \qquad (a)$$

$$(\rho + \rho_1 - 1)\gamma_2 d_2 + (1 - \rho_1)\gamma_1 d_1 \geq 0. \qquad (b)$$

Notice that condition (a) is always satisfied since $\gamma_2 d_2 \geq \gamma_1 d_1$. For condition ($b$), if $\rho + \rho_1 - 1 \geq 0$, then it is always satisfied. If $\rho + \rho_1 - 1 < 0$, then it is satisfied only if $\gamma_2 d_2 / \gamma_1 d_1 \leq (1 - \rho_1)/(1 - \rho - \rho_1)$. ∎

The following corollary follows immediately.

<u>Corollary 3.1:</u> Assuming that Part Type 2 is the part type such that $\gamma_2 d_2 \geq \gamma_1 d_1$. If the setup costs k_1 and k_2 are nonzero, then $\tau_1^* = 0$ and $\tau_2^* = 0$ if and only if

$$0 \leq K \leq \left(T_c^2/2\right)\min\left\{(1 - \rho_2)(\gamma_2 d_2 - \gamma_1 d_1) + \rho\gamma_1 d_1 ; (\rho + \rho_1 - 1)\gamma_2 d_2 + (1 - \rho_1)\gamma_1 d_1\right\}.$$

The following theorem shows that, when the setup costs are zero for both part types, at most one part type will be produced at the demand rate, during the cyclic schedule.

<u>Theorem 3.2:</u> If the setup costs k_1 and k_2 are zero, then at least one of the τ_i's will be equal to zero.

<u>Proof:</u> We prove the theorem by contradiction. Assume that $\tau_1^* > 0$ and $\tau_2^* > 0$, then Q_1^* and Q_2^* are given by (3.10a) and (3.10b), where $K = 0$. In this case, we have

$$Q_1^* = \frac{2q_1 d_2 \gamma_2}{(\alpha_2 d_1 \gamma_1 + \alpha_1 d_2 \gamma_2)};$$

$$Q_2^* = \frac{2q_2 d_1 \gamma_1}{(\alpha_2 d_1 \gamma_1 + \alpha_1 d_2 \gamma_2)}.$$

Using (3.13a) and (3.13b), we get

$$\tau_1^* = \frac{2q_2 \gamma_1 d_1}{d_2(\alpha_2 \gamma_1 d_1 + \alpha_1 \gamma_2 d_2)} - \frac{2q_1 \gamma_2 d_2 \rho_1}{d_1(1 - \rho_1)(\alpha_2 \gamma_1 d_1 + \alpha_1 \gamma_2 d_2)} - \delta > 0;$$

$$\dot{r}_2^* = \frac{2q_1\gamma_2 d_2}{d_1(\alpha_2\gamma_1 d_1 + \alpha_1\gamma_2 d_2)} - \frac{2q_2\gamma_1 d_1\rho_2}{d_2(1-\rho_2)(\alpha_2\gamma_1 d_1 + \alpha_1\gamma_2 d_2)} - \delta > 0.$$

Substituting q_1 and q_2 by their expressions, simplifying and rearranging terms gives:

$$(1-\rho_2)\gamma_1 d_1 - (1+\rho_1)\gamma_2 d_2 > 0 \text{ and } (1-\rho_1)\gamma_2 d_2 - (1+\rho_2)\gamma_1 d_1 > 0.$$

$$\Leftrightarrow \frac{\gamma_2 d_2}{\gamma_1 d_1} < \frac{(1-\rho_2)}{(1+\rho_1)} < 1 \text{ and } \frac{\gamma_2 d_2}{\gamma_1 d_1} > \frac{(1+\rho_2)}{(1-\rho_1)} > 1.$$

Which cannot be. Hence, in the case of no setup costs, $\dot{r}_1^*$ and $\dot{r}_2^*$ cannot be simultaneously nonzero. ∎

Case of zero setup times

The steady state solution for this case can be obtained using the same procedure used for the case where setup times are nonzero. The following theorem shows that in the case of zero setup times, the cyclic schedule must have at least one segment with production at the demand rate (i.e., it is not possible to have the case of Figure 3.3a).

<u>Theorem 3.3:</u> If the setup time is zero for both part types, then there is at least one segment within the cyclic schedule corresponding to a production at the demand rate. That is, if $\delta_1 = \delta_2 = 0$, then it is not possible to have $\dot{r}_1^* = \dot{r}_2^* = 0$.

<u>*Proof:*</u> Assuming that $\dot{r}_1^* = 0$ and $\dot{r}_2^* = 0$, and letting $\delta \to 0$ in the objective function introduced in the proof of Theorem 3.1, gives:

$$F(0,0) = \lim_{\delta \to 0} \frac{\left(K + \frac{H_1 T_0^2}{2} + \frac{H_2 T_0^2}{2}\right)}{T_0} = \frac{K}{0} = \infty$$

Since $T_0 = \delta/(1-\rho)$ and $K>0$.

Hence, this solution cannot be optimal. ∎

In the case of zero setup times, the general structure of the cyclic schedule is as shown in Figure 3.4. The coordinates of the cyclic schedule in this case are given as follows:

$$P1 = \begin{pmatrix} 0 \\ S_2 + s_1(d_2/d_1)\rho_1/(1-\rho_1) \end{pmatrix};$$ (3.15a)

$$P2 = \begin{pmatrix} 0 \\ S_2 + s_1(d_2/d_1)\rho_1/(1-\rho_1) - \tau_1 d_2 \end{pmatrix};$$ (3.15b)

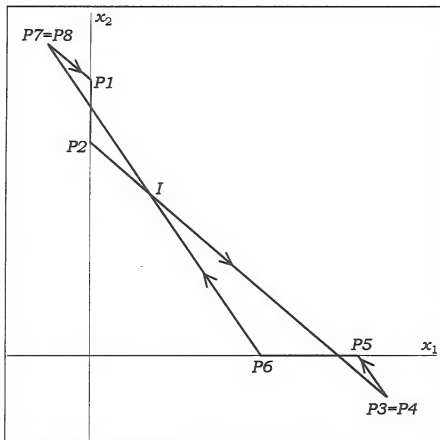$$P3 = P4 = \begin{pmatrix} S_1 \\ s_2 \end{pmatrix};$$ (3.15c)

Figure 3.4: General structure of the cyclic schedule in the case of no setup times.

$$P5 = \begin{pmatrix} S_1 + s_2(d_1/d_2)\rho_2/(1-\rho_2) \\ 0 \end{pmatrix};$$ (3.15d)

$$P6 = \begin{pmatrix} S_1 + s_2(d_1/d_2)\rho_2/(1-\rho_2) - \tau_2 d_1 \\ 0 \end{pmatrix};$$ (3.15e)

$$P7 = P8 = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$$ (3.15f)

where s_i, S_i, Q_i and τ_i $(i=1,2)$ are the optimal values calculated using the same procedure as in the case of nonzero setup times.

3.3 The Multi-Part-Type Case

In this section, we solve the multi-part-type problem numerically. We adopt Zoutendijk's Algorithm (see Bazaraa and Shetty, 1979). To simplify the algorithm, we exploit the structure of the problem and implement the algorithm using closed form expressions. First, we give the mathematical formulation of the multi-part-type problem.

3.3.1 Problem formulation

It is not difficult to extend the two-part-type problem formulation to the m-part-type case. Using the expression of T given by (3.4) and substituting Q_i $(i=1,2)$ given by (3.5) in the objective function given by (3.8) gives the following formulation:

(P)   Minimize $F(\tau) = \dfrac{1}{T}\left(K + \dfrac{1}{2}\sum_{i=1}^{m} H_i(T-\tau_i)^2\right)$

Subject to

$$T = T_0 + \sum_{i=1}^{i=m} \alpha_i \tau_i \, ;$$

$$\tau_i \geq 0, \ i = 1, \, ..., \, m$$

where

$$H_i = \gamma_i d_i (1 - \rho_i), \ i = 1, \ ..., \ m;$$

$$\alpha_i = (1 - \rho_i)/(1 - \rho), \ i = 1, \ ..., \ m;$$

$$T_0 = \delta/(1 - \rho).$$

T_0 is the length of the cyclic schedule when the τ_i's are all zero. Notice that the above formulation is expressed as a function of the τ_i's only. It is not difficult to show that S_i, given by (3.7), is still optimal, which completely characterizes the optimal cyclic schedule. Before we proceed with the algorithm, we propose the following conjecture.

Conjecture 3.1: Based on the strict convexity of the two-part-type objective function, we conjecture that the m-part-type objective function is strictly convex in the τ_i's.

The following proposition guarantees that the optimal τ_i's cannot be infinite.

Proposition 3.2: Problem (P) has always a finite optimal solution.

proof: Assume that the optimal solution of Problem (P) is not finite and label the part types so that τ_1, τ_2, ..., τ_n are infinite and τ_{n+1}, τ_{n+2}, ..., τ_m are finite $(n \le m)$. Without loss of generality, let $\lambda = \tau_1 = \tau_2 = \ ... \ = \tau_n$. Substituting and rearranging terms in the expression of T gives

$$T = \lambda \left(\frac{T_0}{\lambda} + \sum_{i=1}^{i=n} \alpha_i + \sum_{i=n+1}^{i=m} \alpha_i \frac{\tau_i}{\lambda} \right).$$

It is clear that if $\lambda \to \infty$, then τ_1, τ_2, ..., τ_n will all go to infinity. Now, substituting in the objective function, we get

$$F(\tau) = \frac{\left(K + \frac{1}{2}\hat{\lambda}^2\sum_{i=1}^{n}H_i(M_1 + M_2 - 1)^2 + \frac{1}{2}\hat{\lambda}^2\sum_{i=n+1}^{m}H_i\left(M_1 + M_2 - \frac{\tau_i}{\lambda}\right)^2\right)}{\lambda(M_1 + M_2)}$$

where $M_1 = \frac{T_0}{\lambda} + \sum_{i=1}^{i=n}\alpha_i$ and $M_2 = \sum_{i=n+1}^{i=m}\alpha_i\frac{\tau_i}{\lambda}$.

$\lim_{\lambda \to \infty}M_2 = 0$. Hence, $\lim_{\tau \to \infty}F(\tau) = \infty$, which is obviously not optimal since we can always choose a finite feasible solution with a finite objective value. ∎

Since the objective function is strictly convex and the solution is always finite, it follows that Problem (P) has a unique global solution.

3.3.2 Algorithm

First, we state Zoutendijk's algorithm (Bazaraa and Shetty (1979)) for minimizing a differentiable function f in the presence of linear constraints of the form $Ax \leq b$.

Initialization step:

start with a feasible solution x_1. Let $k := 1$, and go to main step.

Main step:

1. given x_k, suppose that A^T and b^T are decomposed into (A_1^T, A_2^T) and (b_1^T, b_2^T) so that $A_1 x_k = b_1$ and $A_2 x_k < b_2$. Let d_k be an optimal solution to the following problem.

(P1) *Minimize* $\nabla f(x_k)^T d$

$\quad$ *Subject to* $\quad A_1 d \leq 0;$

$\quad\quad\quad\quad\quad\quad -1 \leq d_j \leq 1, \ j = 1, \ldots, m,$

if $\nabla f(x_k)^T d_k = 0$, stop; x_k is the optimal solution. Otherwise, go to Step 2.

2. let λ_k an optimal solution to the following line search problem:

(P2) *Minimize* $f(x_k + \lambda d_k)$

    Subject to $0 \le \lambda \le \lambda_{max}$

where

$$\lambda_{max} = \begin{cases} \min\{\hat{b}_i / \hat{d}_i : \hat{d}_i > 0\} & \text{if not all } \hat{d}_i \text{ are negative} \\ \infty & \text{if all } \hat{d}_i \text{ are negative} \end{cases}$$

$$\hat{b} = b_2 - A_2 x_k;$$

$$\hat{d} = A_2 d_k.$$

Let $x_{k+1} = x_k + \lambda_k d_k$, identify the new set of tight constraints at x_{k+1} and update A_1 and A_2 accordingly. Let $k := k+1$, and repeat step 1.

Notice that Problem (P1) is a linear programming problem. Since in our case $A = -I_m$ and $b = 0$, where I_m is the $m \times m$ identity matrix. (P1) can be solved by inspection as follows.

Let $\nabla F_i(\tau)$ be the ith component of the gradient of the objective function $F(\tau)$ in problem (P).

$$\nabla F_i(\tau) = \frac{\left(\alpha_i \sum_{j=1}^{j=m} H_j(T - \tau_j) - H_i(T - \tau_i) - \alpha_i F(\tau) \right)}{T}. \tag{3.16}$$

Let $\mathfrak{I}_k = \{i : \tau_i^k = 0\}$ be the set of indices for which τ_i is equal to zero at iteration k (tight constraints). Then (P1) can be rewritten as follows:

Minimize $\sum_{i=1}^{i=m} \nabla F_i(\tau^k) d_i$

Subject to $0 \le d_i \le 1, \, i \in \mathfrak{I}_k;$

                 $-1 \le d_i \le 1, \, i \notin \mathfrak{I}_k.$

Since for a linear programming problem the optimal solution is always at a vertex, the following procedure gives the optimal solution:

Procedure 2

    IF $i \in \mathfrak{I}_k$ *THEN*

$$IF\ \nabla F_i(\tau^k) > 0\ THEN$$

$$d_i := 0;$$

$$ELSE$$

$$d_i := 1;$$

$$END$$

$$ELSE$$

$$IF\ \nabla F_i(\tau^k) > 0\ THEN$$

$$d_i := -1;$$

$$ELSE$$

$$d_i := 1;$$

$$END$$

$$END$$

The optimization problem (P2) is a line search problem which can be eliminated by calculating the optimal value of λ analytically. Differentiating $F(\tau_k + \lambda d_k)$ with respect to λ and setting the result to zero gives the following quadratic equation in λ.

$$M_1 \sum_{i=1}^{i=m} H_i (M_1 - d_i)^2 \lambda^2\ +\ 2M_2 \sum_{i=1}^{i=m} H_i (M_1 - d_i)^2 \lambda$$
$$+\ 2M_2 \sum_{i=1}^{i=m} H_i (M_1 - d_i)(M_2 - \tau_i^k)\ =\ 0,$$

where $M_1 = \sum_{i=1}^{i=m} \alpha_i d_i$ and $M_2 = T_o + \sum_{i=1}^{i=m} \alpha_i \tau_i^k$.

The positive solution of the above quadratic is given by

$$\lambda^* = -\frac{M_2}{M_1} + \sqrt{\frac{2K + \sum_{i=1}^{m} H_i \left[\frac{M_2}{M_1}(M_1 - d_i) + (M_2 - \tau_i^k) \right]^2}{\sum_{i=1}^{i=m} H_i (M_1 - d_i)^2}}. \tag{3.17}$$

$\hat{b}$ and $\hat{d}$ can be obtained as follows. Since $b = 0$ and $A_2 = -I_n$ (loose constraints), it follows that $\hat{b}_i = \tau_i^k$ and $\hat{d}_i = -d_i^k$ for $i \notin \Im_k$.

The algorithm can now be stated as follows:

Initialization step:

　　start with a feasible solution τ^1. Let $k := 1$, and go to main step.

Main step:

　　1. Identify the set $\Im_k$;

　　　　Apply Procedure 3.2: $\rightarrow d^k$;

　　　　IF $\nabla f(\tau^k)^T d^k = 0$ *THEN*

　　　　　　STOP;

　　　　　　τ^k is the optimal solution.

　　　　ELSE

　　　　　　GOTO Step 2.

　　　　END

　　2. $\hat{b}_i := \tau_i^k$ and $\hat{d}_i := d_i^k$ for $i \notin \Im_k$.

$$\lambda_{max} := \begin{cases} \min\{\hat{b}_i/\hat{d}_i : \hat{d}_i > 0\} & \text{if not all } \hat{d}_i \text{ are negative} \\ \infty & \text{if all } \hat{d}_i \text{ are negative} \end{cases}$$

$$\lambda^k := min\,(\lambda^*, \lambda_{max});$$

$$\tau^{k+1} := \tau^k + \lambda^k d^k;$$

$$k := k + 1;$$

　　　　GOTO Step 1.

3.3.3 Numerical example with backlog allowed

　　　　As an example, consider the ten-product problem proposed by Bomberger (1966), which we extend to the case where backlog is allowed. The backlog cost is 30 times that of inventory holding cost. For this example, we

have normalized the data as follows: the demand rate of each part type is set to 1, the maximum production rate for each part type is divided by its corresponding demand rate, the inventory cost rate of each part type is multiplied by the corresponding demand rate and finally, the backlog cost rate of each part type is multiplied by the corresponding demand rate. The data for the example is as follows.

i	d_i	U_i	δ_i	k_i	c_i^+	c_i^-
1	1	15.3	0.500	130	0.20896	6.2688
2	1	23.5	0.750	200	0.03188	0.9564
3	1	100.0	0.500	110	0.02321	0.6963
4	1	18.8	0.125	10	0.01667	0.5001
5	1	47.5	0.250	30	0.01063	0.3189
6	1	80.0	0.125	20	0.00490	0.1470
7	1	400.0	1.000	310	0.00375	0.1125
8	1	300.0	0.250	50	0.00223	0.0669
9	1	150.0	0.125	5	0.00170	0.0510
10	1	300.0	0.125	5	0.00027	0.0081

The optimal solution is as follows:

i	τ_i^*	t_i^*	S_i^*	s_i^*
1	109.78	1.74	24.1	−0.80
2	0	5.80	126.4	−4.21
3	0	1.36	130.7	−4.36
4	0	7.28	125.0	−4.17
5	0	2.87	129.3	−4.31
6	0	1.71	130.4	−4.35
7	0	0.34	131.7	−4.39
8	0	0.45	131.6	−4.39
9	0	0.91	131.2	−4.37
10	0	0.45	131.6	−4.39

$T^* = 136.46$ *days*

Average cost per day = \$13.05.

3.4 Summary

In this chapter, we have studied the steady state version of the Production and Setup Scheduling Problem. This turned out to be a special Economic Lot Scheduling Problem with controllable production rates and backlog is permitted. We derived the optimal solution of the two-product problem in closed form. For the multi-product problem, we proposed a very simplified version of Zoutendijk's algorithm which requires neither solving a Linear Programming sub-problem for finding a feasible direction, nor performing a line search procedure to determine the next improving solution. Comparison with previous results reported in the literature revealed that savings up to 66% can be obtained when controllable production rates are allowed.

CHAPTER 4
OPTIMAL TRANSIENT SOLUTION OF A TWO-PART-TYPE
MANUFACTURING SYSTEM WITH SETUP TIMES AND NO SETUP COSTS
DETERMINISTIC SYSTEM

4.1 Introduction

In this chapter, we study the transient behavior of the deterministic two-part-type system. We consider the case where setup costs are negligible and the system is heavily loaded. In Section 4.2, we present the optimal location of the cyclic schedule in x-space. In Section 4.3, we show how we partition the x-space into two mutually exclusive major regions $\Re^u$ and $\Re^o$. In Section 4.4, we determine the optimal transient solution in Region $\Re^o$. In Section 4.5, we determine the optimal transient solution in Region $\Re^u$, using an algorithmic approach. In Section 4.6, we analytically determine the optimal transient solution in Region $\Re^u$. We conclude this chapter with a summary of results.

4.2 Optimal Location of the Cyclic Schedule

We assume that the setup costs k_1 and k_2 are zero for both part types and that the conditions of Theorem 3.1 of Chapter 3 are fully satisfied. This guarantees that the optimal cyclic schedule has the shape shown in Figure 4.1. That is, there is no production at the demand rate at the axis. In this case, The location of the cyclic schedule is obtained by minimizing the average inventory and backlog costs incurred during one complete cycle and is given by the following points in x-space:
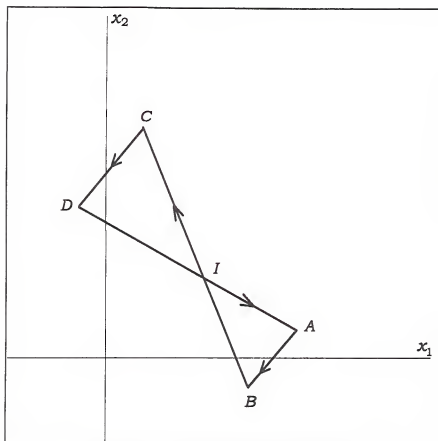
Figure 4.1: Optimal cyclic schedule

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \ B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \ C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}, \ D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}, \ I = \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}, \tag{4.1}$$

where

$$A_1 = \frac{c_1^-}{(c_1^+ + c_1^-)} \frac{(1 - \rho_1)}{(1 - \rho)} \delta d_1, \qquad A_2 = \delta_2 d_2 - \frac{c_2^+}{(c_2^+ + c_2^-)} \frac{(1 - \rho_2)}{(1 - \rho)} \delta d_2; \tag{4.2}$$

$$B_1 = \frac{c_1^-}{(c_1^+ + c_1^-)} \frac{(1 - \rho_1)}{(1 - \rho)} \delta d_1 - \delta_2 d_1, \qquad B_2 = -\frac{c_2^+}{(c_2^+ + c_2^-)} \frac{(1 - \rho_2)}{(1 - \rho)} \delta d_2; \tag{4.3}$$

$$C_1 = \delta_1 d_1 - \frac{c_1^+}{(c_1^+ + c_1^-)} \frac{(1 - \rho_1)}{(1 - \rho)} \delta d_1, \qquad C_2 = \frac{c_2^-}{(c_2^+ + c_2^-)} \frac{(1 - \rho_2)}{(1 - \rho)} \delta d_2; \tag{4.4}$$

$$D_1 = -\frac{c_1^+}{(c_1^+ + c_1^-)} \frac{(1 - \rho_1)}{(1 - \rho)} \delta d_1, \qquad D_2 = \frac{c_2^-}{(c_2^+ + c_2^-)} \frac{(1 - \rho_2)}{(1 - \rho)} \delta d_2 - \delta_1 d_2; \tag{4.5}$$

Here and elsewhere in the dissertation, $\rho_i = d_i/U_i$ $(i = 1,2)$ denotes the utilization factor of the machine by Type i parts, and $\rho = \rho_1 + \rho_2$ is the utilization factor of the machine. For the problem to be feasible, we need to have $\rho < 1$. $\delta = \delta_1 + \delta_2$ is the sum of the setup times.

The total average cost over one cyclic schedule is given by

$$Average\ Cost = \frac{1}{2} \frac{c_1^- c_1^+}{(c_1^- + c_1^+)} \frac{(1-\rho_1)}{(1-\rho)} d_1 \delta + \frac{1}{2} \frac{c_2^- c_2^+}{(c_2^- + c_2^+)} \frac{(1-\rho_2)}{(1-\rho)} d_2 \delta.$$

4.3 Partition of the x-Space into Two Mutually Exclusive Regions

Based on Definitions 2.3 and 2.4 (Chapter 2), the optimal transient solution corresponds to a state trajectory emanating in x-space and reaching the cyclic schedule with minimum total cost and in finite time. The problem consists in finding such optimal trajectories for every initial production surplus level in x-space. Before we start deriving the optimal transient solution, we notice the following facts.

Fact 4.1: Since the cost coefficients in the HJB equation of Theorem 2.1 are function of x, given an optimal choice $\sigma^*(\tau)$, $t \leq \tau \leq t_s$, the x-space can be partitioned into mutually exclusive regions (Gershwin (1993)), each corresponding to an optimal setup state $\sigma^* = (\sigma_1^*, \sigma_2^*, \sigma_{12}^*, \sigma_{21}^*)$ and an optimal production rate $u^* = (u_1^*, u_2^*)$ which belongs to the set Ω^*, defined in Chapter 2.

Another important fact drawn from the HJB equation is given as follows:

Fact 4.2: given an optimal choice $\sigma^*(\tau)$, $t \leq \tau \leq t_s$, J^τ, the transient cost component does not depend explicitly on time and is piecewise quadratic. Therefore, the cost coefficients of the HJB equation in Theorem 2.1 are linear in

x. This implies that the boundaries of the regions in *x*-space must be linear (Gershwin, 1993).

For the subsequent analysis, the following important quantities need to be defined. First we need to find analytic expressions for the equations of line (*A*,*D*) (L1), line (*B*,*C*) (L2), line parallel to (*A*,*D*) containing Point *C* (L21) and the line parallel to (*B*,*C*) containing Point *A* (L12) (Figure 4.2). These line equations are easy to calculate since we know the coordinates of the points *A*, *B*, *C*, and *D* (from the steady state optimal solution).

$$\text{Line} \quad \text{L1:} \quad d_2\rho_1(x_1 - A_1) + d_1(1 - \rho_1)(x_2 - A_2) = 0; \tag{4.6}$$

$$\text{Line} \quad \text{L2:} \quad d_2(1 - \rho_2)(x_1 - C_1) + d_1\rho_2(x_2 - C_2) = 0. \tag{4.7}$$

$$\text{Line L12:} \quad d_2(1 - \rho_2)(x_1 - A_1) + d_1\rho_2(x_2 - A_2) = 0; \tag{4.8}$$

$$\text{Line L21:} \quad d_2\rho_1(x_1 - C_1) + d_1(1 - \rho_1)(x_2 - C_2) = 0; \tag{4.9}$$

Having the equations of Lines L1 and L2, we can easily find the coordinates of the intersection point *I*, which is also of importance in the subsequent analysis. Intersection Point *I* is given by

$$I_1 = \frac{(1 - \rho_1)}{(1 - \rho)}\left(\frac{c_1^-}{c_1^+ + c_1^-}d_1\delta - d_1\delta_2\right), \qquad I_2 = \frac{(1 - \rho_2)}{(1 - \rho)}\left(\frac{c_2^-}{c_2^+ + c_2^-}d_2\delta - d_2\delta_1\right). \tag{4.10}$$

Notice that intersection Point *I* can be located in any quadrant of the production surplus space, depending on the values of the penalty function coefficients and the setup time duration.

Without loss of generality we index the parts such that Part Type 1 is the part type with the larger setup time (i.e., $\delta_1 \geq \delta_2$).

To obtain the optimal transient solution, we divide the surplus space into two mutually exclusive major regions: $\Re^u$ and $\Re^o$.
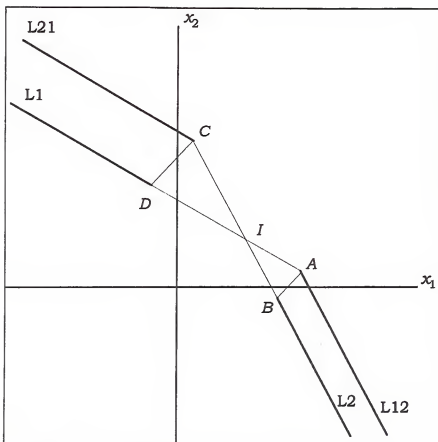
Figure 4.2: Illustration of lines L1, L2, L12 and L21.

$\mathfrak{R}^u$ is the region in x-space located below Lines L12 and L21. $\mathfrak{R}^o$ is the region in x-space located above Lines L12 and L21 (Figure 4.3). Algebraically,
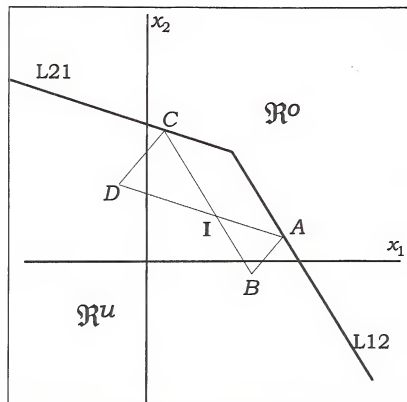
$$\mathfrak{R}^u = \{(x_1, x_2) \mid d_2(1-\rho_2)(x_1 - A_1) + d_1\rho_2(x_2 - A_2) < 0;$$
$$d_2\rho_1(x_1 - C_1) + d_1(1-\rho_1)(x_2 - C_2) < 0\}.$$
$$\mathfrak{R}^o = \{(x_1, x_2) \mid d_2(1-\rho_2)(x_1 - A_1) + d_1\rho_2(x_2 - A_2) \geq 0$$
$$d_2\rho_1(x_1 - C_1) + d_1(1-\rho_1)(x_2 - C_2) \geq 0\}.$$

We divide the analysis into two parts. In the first part, we determine the optimal transient solution for all initial surplus levels in Region $\mathfrak{R}^o$. In the second part, we do the same for all initial surplus levels in Region $\mathfrak{R}^u$.

Figure 4.3: Regions $\mathfrak{R}^u$ and $\mathfrak{R}^o$.

4.4 Optimal Transient Solution in Region $\mathfrak{R}^o$

As mentioned above, our task is to bring the initial surplus levels in x-space to a point on the cyclic schedule. Because, once on the cyclic schedule, the production surplus will behave according to it thereafter (since the machine is reliable), and we know that operating according to the cyclic schedule is optimal (from the steady state solution). The solution approach for initial surplus levels in Region $\mathfrak{R}^o$ is obtained by inspection. However, it is easy to see that the solution is indeed optimal. For instance, from Theorem 2.1 and Fact 4.1, we know that the optimal production rates are piecewise constant and are selected from a very small set. Hence, from the system dynamics (differential equations (2.1)), it is easy to verify that the optimal trajectories in x-space are

piecewise linear and therefore it would be not difficult to see that they are indeed optimal.

We will provide a detailed solution for the case where Intersection Point I is located in the first quadrant (Figure 4.3). The positive location of Intersection Point I may be the most common in practice. It has been reported in the literature that the backlog and inventory costs, c_i^- and c_i^+ ($i = 1, 2$), are usually chosen at least in the ratio of 5 to 1 (because backlog usually results in sales and customer's good will losses and other undesirable effects). The setup times δ_i ($i = 1, 2$) are of the same order of magnitude. Hence, from the expression of the coordinates of Intersection Point I in (4.10), we have:

$$c_i^- / (c_i^- + c_i^+) > \delta_j / (\delta_1 + \delta_2) \quad \text{for } i = 1, 2;\ j = 1, 2 \text{ and } j \neq i.$$

Hence, i_1 and i_2 are positive numbers. In any case, the solution procedure is exactly the same for Intersection Point I located in any other quadrant in x-space.

Throughout the dissertation, we assume that initially the machine is not setup to either part type. For initial surplus levels in Region $\Re^\circ$, we can state our problem as follows: *Subject to constraints (2.1)-(2.6), find an optimal trajectory (i.e., with minimum cost) emanating from an initial point $x(t) \in \Re^\circ$ in x-space and reaching the cyclic schedule at a point $x(t_s)$, where it would be possible to move according to it thereafter.*

The solution to the above problem depends on the initial surplus levels. Therefore, we further partition Region $\Re^\circ$ into mutually exclusive regions G, G1, and G2 by introducing two linear boundaries LG1 and LG2. Line LG1 (LG2) is the line for which the surplus level of Part Type 1 (Part Type 2) is exactly equal to $d_1\delta_1$ ($d_2\delta_2$). This partition is shown in Figure 4.4. Its justification will be apparent from the solution below.

4.4.1 Initial surplus levels in Region G

In this case, both surplus levels are positive. Therefore, the production is ahead of the demand for both parts and it is optimal to let the surplus levels deplete at the maximum possible rate. That is, we set the production rates to zero. However, we also want to reach the cyclic schedule as quickly as possible. Hence, we need to decide which setup we must prepare the machine for, and when is the best time to do it. To answer these questions, we further partition Region G into six regions: G11, G12, G21, G22, H1 and H2 located in the first quadrant (Figure 4.5).
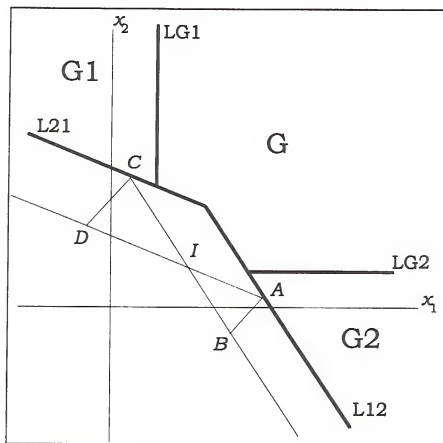

Figure 4.4: Further partition of Region $\Re^\circ$.

containing Point I with direction (d_1, d_2). The optimal trajectory is obtained by setting the production rates to zero. This corresponds to a trajectory moving downward in the southwest direction with speed $(-d_1, -d_2)$. When the trajectory touches the boundary of Region G12 on Line L12, we immediately start a setup for Part Type 2. At the end of the setup change the trajectory touches the cyclic schedule at a point on the Segment $[I, C]$ (trajectory τ_{12} in Figure 4.5).

Initial surplus in Region G21

Region G21 is defined as follows:

$$G21 = \{(x_1, x_2) | -d_2(x_1 - A_1) + d_1(x_2 - A_2) \geq 0;$$
$$d_2(x_1 - I_1) - d_1(x_2 - I_2) > 0;$$
$$d_2\rho_1(x_1 - C_1) + d_1(1 - \rho_1)(x_2 - C_2) \geq 0\},$$

where the first constraint corresponds to the half space above Line LA. The second constraint corresponds to the half space below Line LI. The third constraint corresponds to the half space above Line L21. LA is the line containing Point A with direction (d_1, d_2). In this case, the optimal trajectory is obtained by first setting the production rates to zero. This corresponds to a trajectory moving downward in the southwest direction with speed $(-d_1, -d_2)$. When the trajectory touches the boundary of Region G21 on Line L21, we immediately start a setup for Part Type 1. At the end of the setup, the trajectory touches the cyclic schedule at a point on the Segment $[I, A]$ (trajectory τ_{21} in Figure 4.5).

Initial surplus in Region G22

Region G22 is defined as follows:

$$G22 = \{(x_1, x_2) | x_2 - \delta_2 d_2 \geq 0; \ d_2(x_1 - A_1) - d_1(x_2 - A_2) > 0\},$$
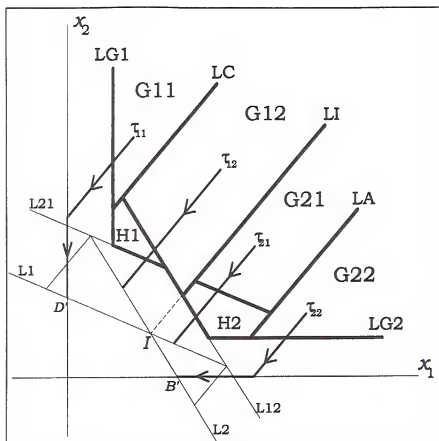
Figure 4.5: Optimal trajectories emanating at points
in Regions G11, G12, G21, G22, H1 and H2.

where the first constraint corresponds to the half space above Line LG2. The
second constraint correspond to the half space below Line LA. In this case, the
optimal trajectory is obtained by first setting the production rates to zero. This
corresponds to a trajectory moving downward in the southwest direction with
speed $(-d_1, -d_2)$. When the surplus of Part Type 2 reaches level $\delta_2 d_2$ (that is the
trajectory hits Line LG2), we immediately start a setup for Part Type 2. At the
end of the setup, the surplus level of Part Type 2 is exactly zero, that of Part
Type 1 is still positive and we still have not reached the cyclic schedule.
Therefore, we need to decrease the surplus level of Part Type 1 and keep the
surplus level of Part Type 2 at the zero level until we touch the cyclic schedule.
To do this, we produce Part Type 2 at the demand rate d_2. This corresponds to

a trajectory moving along the x_1 axis with speed $(-d_1, 0)$, until the cyclic schedule is touched at Point B' (trajectory τ_{22} in Figure 4.5).

<u>Initial surplus levels in Region H1</u>

Region H1 is defined by

$$H1 = \{(x_1, x_2)| \ x_1 - \delta_1 d_1 \geq 0;$$
$$d_2(x_1 - C_1) - d_1(x_2 - C_2) > 0;$$
$$d_2\rho_1(x_1 - C_1) + d_1(1 - \rho_1)(x_2 - C_2) \geq 0;$$
$$-d_2(1 - \rho_2)(x_1 - A_1) - d_1\rho_2(x_2 - A_2) \geq 0\}.$$

The first constraint of this set corresponds to the half space below Line LC. The second constraint corresponds to the half space to the right of Line LG1. The third constraint corresponds to the half space above Line L21 and the fourth constraint corresponds to the half space below Line L12. Region H1 can be further partitioned into two mutually exclusive regions: H11 and H12. Region H11 is defined by

$$H11 = \{(x_1, x_2)| \ x_1 - \delta_1 d_1 \geq 0$$
$$d_2(x_1 - C_1) - d_1(x_2 - C_2) > 0;$$
$$-d_2(x_1 - g_{11}) + d_1(x_2 - g_{12}) \geq 0;$$
$$-d_2(1 - \rho_2)(x_1 - A_1) - d_1\rho_2(x_2 - A_2) \geq 0\},$$

where $g_1 = (g_{11}, g_{12})$ is the point in x-space given by

$$g_{11} = \delta_1 d_1 \text{ and } g_{12} = C_2 + (C_1 - \delta_1 d_1)d_2\rho_1/d_1(1 - \rho_1).$$

Point g_1 is the intersection of Lines L21 and Lg1 (Figure 4.6). For initial surplus levels in Region H11, the optimal trajectories are obtained in the same manner as for initial surplus levels in Region G11 (trajectory τ_{11} in Figure 4.6).

Region H12 is defined by

$$H12 = \{(x_1, x_2)| \ d_2\rho_1(x_1 - C_1) - d_1(1 - \rho_1)(x_2 - C_2) > 0;$$

$$-d_2(x_1 - g_{11}) + d_1(x_2 - g_{12}) \geq 0;$$
$$-d_2(1 - \rho_2)(x_1 - A_1) - d_1\rho_2(x_2 - A_2) \geq 0\}.$$

For initial surplus levels in Region H12, the optimal trajectories are obtained in the same manner as for initial surplus levels in Region G21 (trajectory τ_{12} in Figure 4.6).

<u>Initial surplus levels in Region H2</u>

Region H2 is defined by

$$H2 = \{(x_1, x_2) | \ x_2 - \delta_2 d_2 \geq 0;$$
$$d_2(x_1 - I_1) - d_1(x_2 - I_2) > 0;$$
$$d_2(1 - \rho_2)(x_1 - A_1) - d_1\rho_2(x_2 - A_2) \geq 0;$$
$$-d_2\rho_1(x_1 - C_1) - d_1(1 - \rho_1)(x_2 - C_2) > 0;$$
$$-d_2(x_1 - A_1) + d_1(x_2 - A_2) > 0\}.$$

The first constraint of this set corresponds to the half space above Line L12, the second constraint corresponds to the half space above Line LG2, the third constraint corresponds to the half space below Line LI, the fourth constraint corresponds to the half space below Line L21, and the fifth constraint corresponds to the half space above Line LA. Region H2 can be further partitioned into two mutually exclusive regions: H21 and H22.

Region H21 is defined by

$$H21 = \{(x_1, x_2) | d_2(x_1 - I_1) - d_1(x_2 - I_2) > 0;$$
$$d_2(1 - \rho_2)(x_1 - A_1) - d_1\rho_2(x_2 - A_2) \geq 0;$$
$$-d_2\rho_1(x_1 - C_1) - d_1(1 - \rho_1)(x_2 - C_2) > 0;$$
$$-d_2(x_1 - g_{21}) + d_1(x_2 - g_{22}) \geq 0\},$$

where $g_2 = (g_{21}, g_{22})$ is the point in x-space given by
$$g_{21} = A_1 + (A_2 - \delta_2 d_2)d_1\rho_2/d_2(1 - \rho_2) \text{ and } g_{22} = \delta_2 d_2.$$

Point g_2 is the intersection of Lines L12 and Lg2 (Figure 4.6). For initial surplus levels in Region H21, the optimal trajectories are obtained in the same manner as for initial surplus levels in Region G12 (trajectory τ_{12} in Figure 4.6).

Region H22 is defined by

$$H22 = \{(x_1, x_2) | x_2 - \delta_2 d_2 \geq 0;$$
$$-d_2(x_1 - A_1) - d_1(x_2 - A_2) > 0;$$
$$-d_2 \rho_1(x_1 - C_1) - d_1(1 - \rho_1)(x_2 - C_2) > 0;$$
$$-d_2(x_1 - g_{21}) + d_1(x_2 - g_{22}) > 0\}.$$

For initial surplus levels in Region H22, the optimal trajectories are obtained in the same manner as for initial surplus levels in Region G22 (trajectory τ_{22} in Figure 4.6).
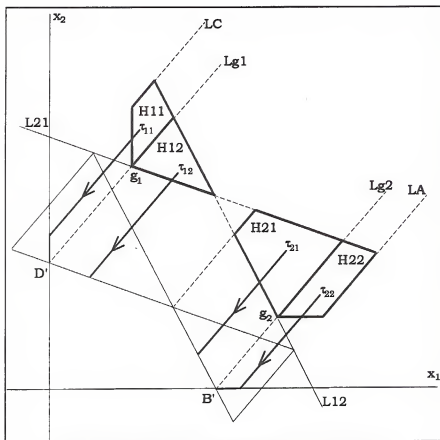

Figure 4.6: Partition of Regions H1 and H2.

4.4.2 Initial surplus levels in Region G1

Region G1 is defined by

$$G1 = \{(x_1, x_2) | -x_1 + \delta_1 d_1 > 0; \ d_2 \rho_1 (x_1 - C_1) + d_1 (1 - \rho_1)(x_2 - C_2) \geq 0\}.$$

The first constraint of this set corresponds to the half space to the left of Line LG1 (Figure 4.7). The second constraint corresponds to the half space above Line L21. Notice that in this region, the surplus level of Part Type 2 is always positive and the surplus level of Part Type 1 is less than $\delta_1 d_1$. Therefore, we need to produce Part Type 1 so that we can reach the cyclic schedule. Suppose that the surplus level of Part Type 1 is positive. In this case even if we started a setup for Part Type 1, we would endup with a backlog of Part Type 1, since during the setup for Part Type 1, $\delta_1 d_1$ amount of Part Type 1 is depleted. The optimal trajectories emanating from Region G1 and leading to the cyclic schedule are obtained as follows: First we start a setup for Part Type 1. This corresponds to a trajectory moving downward with speed $(-d_1, -d_2)$. When the setup is completed, x_1 is negative and x_2 is positive. Hence, we produce Part Type 1 at the maximum machine production rate to eliminate the backlog of Part Type 1. This corresponds to a trajectory moving southeast with speed $(U_1 - d_1, -d_2)$. When the surplus level of Part Type 1 becomes zero, that of Part Type 2 is still positive (Point where the trajectory hits the x_2 axis), and we still have not reached the cyclic schedule. Therefore, we need to decrease the surplus level of Part Type 2 and keep the surplus level of Part Type 1 at the zero level until we touch the cyclic schedule. The optimal way to do this is to produce Part Type 1 at the demand rate d_1. This corresponds to a trajectory moving downward along the x_2 axis with speed $(0, -d_2)$ until the cyclic schedule is touched at Point D' (trajectories τ_1 and τ_1' in Figure 4.7).

4.4.3 Initial surplus levels in Region G2

Region G2 is defined by

$$G2 = \{(x_1, x_2) | -x_2 + \delta_2 d_2 > 0; d_2(1 - \rho_2)(x_1 - A_1) + d_1 \rho_2 (x_2 - A_2) \geq 0\}.$$

The first constraint of this set corresponds to the half space above Line LG2 (Figure 4.7). The second constraint corresponds to the half space above Line L12. In this region, the surplus level of Part Type 1 is always positive and the surplus level of Part Type 2 is less than $\delta_2 d_2$. This is the symmetric case of initial surplus levels in Region G1. That is, the optimal way to get to the cyclic schedule is to setup the machine for Part Type 2 first, produce Part Type 2 at the maximum machine production rate until its surplus level becomes zero. At that point change the level of production of Part Type 2 to d_2 and continue producing this part until the trajectory touches the cyclic schedule at Point B'. (trajectories τ_2 and τ_2' in Figure 4.7).

To summarize the control actions in Region $\mathfrak{R}^\circ$, let $x = (a, b)$ be the vector of initial surplus levels. Then, we have

• If $x \in$ G1:

    1: Setup the machine for Part Type 1;

    2: Produce Part Type 1 at the rate U_1 until the surplus level of Part Type 1 becomes 0;

    3: Change the production rate to d_1;

    4: When the surplus level of Part Type 2 becomes $x_2 = A_2 + A_1 d_2 \rho_1 / d_1 (1 - \rho_1)$, switch to the control actions of the cyclic schedule.

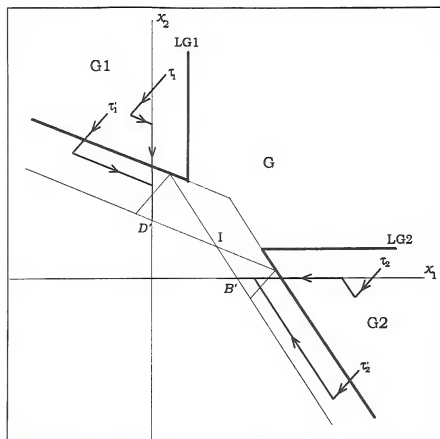• If $x \in$ G11$\bigcup$H11:

    1: Do not produce either part type;

Figure 4.7: Optimal trajectories emanating from Regions G1 and G2.

2: When the surplus level of Part Type 1 becomes $\delta_1 d_1$, start a setup for Part Type 1;

3: Produce Part Type 1 at the demand rate d_1;

4: When the surplus level of Part Type 2 becomes $x_2 = A_2 + A_1 d_2 \rho_1 / d_1 (1 - \rho_1)$, switch to the control actions of the cyclic schedule.

• If $x \in G12 \cup H21$:

1: Do not produce either part type;

2: When the surplus level of Part Type 2 reaches Level l_2, immediately start a setup for Part Type 2. l_2 is given by $l_2 = (1 - \rho_2)(b - (d_2/d_1)a) + (d_2/d_1)(1 - \rho_2)A_1 + \rho_2 A_2$;

   3: When the setup is over, switch to the cyclic schedule
      control actions.

- If $x \in G21 \cup H12$:

   1: Do not produce either part type;
   2: When the surplus level of Part Type 1 reaches Level l_1,
      immediately start a setup for Part Type 1. l_1 is given   by
      $$l_1 = \rho_1 (b - (d_2/d_1)a) + (d_2/d_1)\rho_1 C_1 + (1 - \rho_1)C_2;$$
   3: when the setup is over, switch to the cyclic schedule
      control actions.

- If $x \in G22 \cup H22$:

   1: Do not produce either part type;
   2: When the surplus level of Part Type 2 becomes $\delta_2 d_2$, start a
      setup for Part Type 2;
   3: when the setup is over, produce Part Type 2 at the demand
      rate d_2;
   4: When the surplus level of Part Type 1 becomes
      $x_1 = C_1 + C_2 d_1 \rho_2 / d_2 (1 - \rho_2)$, switch to the control actions of the
      cyclic schedule.

- If $x \in G2$:

   1: Setup the machine for Part Type 2;
   2: when the setup is over, produce Part Type 2 at the rate U_2;
   3: When the surplus level of Part Type 2 becomes 0, change the
      production rate to d_2;
   4: When the surplus level of Part Type 1 becomes
      $x_1 = C_1 + C_2 d_1 \rho_2 / d_2 (1 - \rho_2)$, switch to the control actions of the
      cyclic schedule.

Notice that all optimal trajectories, emanating in Region $\Re^\circ$, reach the cyclic schedule by moving downward and that only one setup for either part type is performed before the cyclic schedule is reached. In other words, for initial surplus levels in Region $\Re^\circ$, the cyclic schedule is always reached from above with only one setup.

4.5 Optimal Transient Solution in Region $\Re^u$ (Algorithmic Approach)

In this section, we develop an algorithm to obtain the optimal trajectories emanating in Region $\Re^u$ and reaching the cyclic schedule in finite time. Before we proceed with the algorithm, we establish the following facts and definitions.

In the previous subsection, we showed that for initial surplus levels in Region $\Re^\circ$, the cyclic schedule is always reached in finite time from above with only one setup. For initial surplus levels in Region $\Re^u$, if we immediately started a setup for a part type, at the end of the setup, we would miss either Segment $[A,D]$ if we set up for Part Type 1, or Segment $[B,C]$ if we set up for Part Type 2 (which are the target segments for the surplus to stay on the cyclic schedule). To bring the surplus levels to either segments in finite time, we need to generate a surplus excess for the part type the machine is set up for, so that when we switch to the production of the other part type, we endup on or above the appropriate segment of the cyclic schedule. After the setup, we need to produce the part type we set up the machine for. Based on Theorem 2.1, we can set the production rate to zero, to the demand rate, or to the maximum rate. It is clear that if we set the production rate to zero, both surplus levels deplete and the generated trajectory moves in the direction $(-d_1, -d_2)$ and hence further drifts away from the cyclic schedule (Figure 4.8). If we produced at the

demand rate the part type the machine is set up for, we would keep the surplus level of this part type constant, while the surplus of the other part type depletes. In this case, the generated trajectory will move parallel to one of the axis in x-space in the direction that further drifts away from the cyclic schedule (Figure 4.8). If we produced at the maximum production rate the part type the machine is set up for, the surplus level of this part type increases, that of the other part type decreases. This way it is possible to generate an excess surplus for the part type being produced. The following fact formally states this result.

Fact 4.3: For initial surplus levels in Region $\mathfrak{R}^u$, the only way to progress toward the cyclic schedule is by producing at the maximum production rate whenever it is possible.

Based on Fact 4.3, the trajectories emanating in Region $\mathfrak{R}^u$ either move along direction $(-d_1, -d_2)$ during a setup, along direction $(U_1 - d_1, -d_2)$ during the production of Part Type 1 or along direction $(-d_1, U_2 - d_2)$ during the production of Part Type 2. Therefore, the cyclic schedule cannot be reached from below (i.e., from points in Region $\mathfrak{R}^u$) in finite time. To be able to reach the cyclic schedule in finite time, starting in Region $\mathfrak{R}^u$, we should bring the surplus levels in Region $\mathfrak{R}^o$, then apply the optimal control actions of that Region (since we know that those controls lead to the cyclic schedule in finite time). Doing so will generate a surplus excess of either part type. To reach the cyclic schedule with the least amount of excess, we must bring the surplus levels to the boundary of Region $\mathfrak{R}^o$ (Line Lij $i = 1,2$; $j = 1,2$; $i \neq j$) with minimum cost. Once on the boundary Lij, we switch to Part Type j and produce this part type (this corresponds to a trajectory moving along Line Lj) until the cyclic schedule is reached at Point D if $j=1$ or Point B if $j=2$ (Figure 4.8). The following fact formally states this result.
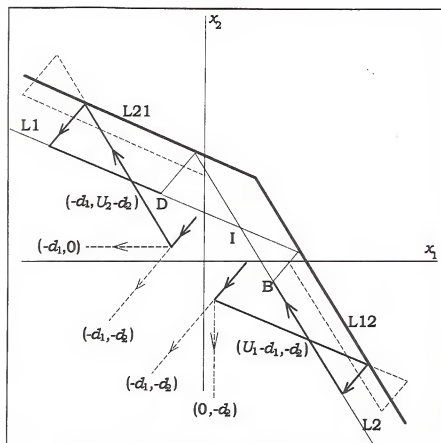
Figure 4.8: Illustration of Facts 4.3 and 4.4.

Fact 4.4: To reach the cyclic schedule in finite time, starting with initial surplus levels in Region $\mathfrak{R}^u$, the trajectory leading to the cyclic schedule must touch the boundary Lij of Region $\mathfrak{R}^c$ just before switching to Part Type j and reaching the cyclic schedule at one of the points B or D on the cyclic schedule.

Definition 4.1: We say that a trajectory is following Direction D_i, if it moves parallel to Line Li (Figure 4.2) in the direction of increasing x_1. That is, the machine is producing Part Type i ($i = 1, 2$).

Remark: Since for surplus levels in Region $\mathfrak{R}^u$ the machine always produces at its maximum rate, the trajectories will move either along Direction D_1, along Direction D_2 or along Direction D_0 (during a setup for a part type).

Definition 4.2: We call a $D_i - n - step$ trajectory $(i = 1, 2; n > 1)$, a trajectory that performs alternately m_i setups-production runs of Part Type i and m_j setups-production runs of Part Type j $(j \neq i)$, with the initial setup for Part Type i and the last segment touching the cyclic schedule at Point B or Point D. If n is even then, $m_i = m_j = n/2$. If n is odd then, $m_i = (n+1)/2$ and $m_j = (n-1)/2$. Figure 4.9 shows a $D_1 - 2 - step$ and a $D_2 - 2 - step$ trajectories.

We can now state the problem as follows: *Given surplus levels in Region $\Re^u$, find a $D_{i^*} - n^* - step$ trajectory that minimizes the cost of reaching the cyclic schedule and satisfying constraints (2.1)-(2.6). Where, n^* is the optimal number of steps (setups-production runs) before reaching the cyclic schedule and D_{i^*} is the first direction to follow (i.e., the initial setup).*
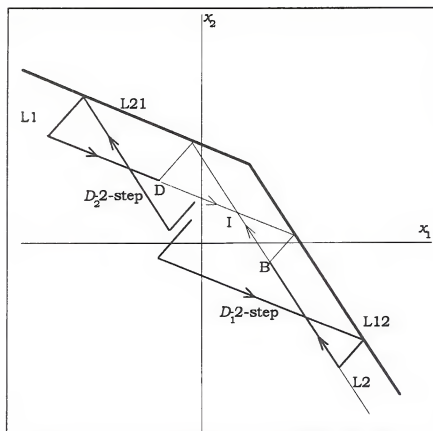

Figure 4.9. $D_1 - 2 - step$ and $D_2 - 2 - step$ trajectories.

As a consequence of Facts 4.3 and 4.4, the optimal trajectory leading to the cyclic schedule in finite time is at least $D_i - 2 - step$. Since, we need a setup change at the initial point and one more, once the boundary of Region $\Re^\circ$ is reached. This leads to the following proposition.

Proposition 4.1: Let $C_i^n(x)$ ($i = 1,2$) be the cost of moving along a $D_i - n - step$ trajectory with initial surplus point x in surplus space. An upper bound on the optimal cost of reaching the cyclic schedule is given by $\min_{i=1,2}\{C_i^2(x)\}$.

Proof: Assume that the optimal trajectory with initial point x is $D_{i^*} - n^* - step$ ($n^* \geq 2$). We know that an optimal trajectory originating at Point x in surplus space is at least $D_i - 2 - step$. It follows that,

$C_{i^*}^{n^*}(x) \leq C_1^2(x)$ and $C_{i^*}^{n^*}(x) \leq C_2^2(x)$. Therefore, $C_{i^*}^{n^*}(x) \leq \min_{i=1,2}\{C_i^2(x)\}$. ■

Based on Facts 4.3 and 4.4, The optimal trajectory emanating at a point in Region $\Re^u$ and leading to the cyclic schedule in finite time can be obtained as follows: Given an initial surplus point in Region $\Re^u$, we choose the first setup and calculate the cost of the trajectory leading to the cyclic schedule with two setups only. At this point, we have a $D_i - 2 - step$ trajectory. Where, i is the initial setup for Part Type i. The next step is to try to lower the cost of the current trajectory by introducing a setup for the other part type so that the cyclic schedule is reached at the opposite side. If the cost can be reduced, the obtained new trajectory is a $D_i - 3 - step$ trajectory. We keep trying to reduce the cost of the current trajectory by introducing, each time, a setup before the cyclic schedule is reached until we cannot lower the cost anymore. The resulting trajectory is an optimal $D_i - n - step$ trajectory (provided we start with a setup for Part Type i) emanating in Region $\Re^u$ and reaching the cyclic

schedule in finite time. In the same manner, we obtain the optimal $D_j - n - step$ ($j{\neq}i$) trajectory starting with a setup for Part type j first. The optimal trajectory would be the one with the lower cost.

The numerical solution of various examples suggests that the above procedure be further simplified based on the following conjecture.

Conjecture 4.1: The initial setup of the optimal trajectory is given by
$$i^* = \arg\min_{i=1,2}\{C_i^2(x)\}.$$

Conjecture 4.1 implies that we only need to compare two $D_i - 2 - step$ trajectories instead of comparing two $D_i - n - step$ trajectories. Which is simpler and faster to compute. In other words, the initial setup is optimal and once we obtain the initial setup, we do not need to go back and check whether starting with the other setup is optimal. Although we do not have a mathematical proof for this observation, the numerical solution of an extensive number of cases agree with Conjecture 4.1.

Conjecture 4.1 along with the above procedure lead to the following algorithm which is called Direction Sweeping Algorithm (DSA).

Direction Sweeping Algorithm

Notation:

$x(0)$: Initial surplus in x-space (Input);

$\{x(0),x(1), ..., x(k)\}$: Optimal trajectory (Output);

J_x: Optimal trajectory cost (Output);

$C(x(k))$: Cost of the trajectory up to the point $x(k)$;

$Cgm(Y,Z)$: Cost along Segment $[Y,Z]$.

Dr: Direction of search. $Dr = D_1$ or $Dr = D_2$ as defined in Definition 4.3.

$\overline{Dr} = 1$ if $Dr = 2$ and $\overline{Dr} = 2$ if $Dr = 1$.

$C_{Dr}^n(X)$: Cost of Dr-n-step trajectory starting at the point X in x-space.

$I(X,Dr)$: Intersection point of the line containing X and with direction Dr, with Line $L_{Dr,\overline{Dr}}$, where $L_{Dr,\overline{Dr}} = L_{12}$ or $L_{Dr,\overline{Dr}} = L_{21}$, depending on the direction of search Dr.

$P_{Dr} : P_1 = B$ and $P_2 = D$.

Main Algorithm:

// Initialization:

$k := 1;$

$m := 1;$

$STOP := FALSE;$

// Find an upper bound on the cost of the optimal trajectory emanating from x(0):

$(J_x, Dr) := \min_{Dr=1,2}\{C_{Dr}^2(x(0))\};$

// Update trajectory:

$k := k+1;$

$x(k) := x(k-1) + \delta_{Dr}(-d_1, -d_2);$

$\overline{Dr} := not\ Dr;$

// Update trajectory cost up to the point x(k):

$C(x(k)) := Cgm(x(k-1), x(k));$

WHILE $STOP \neq TRUE$ DO

    $swap(Dr, \overline{Dr});$

    // Add a new step to the trajectory:

    $(C_{Dr}^2(Y(m)), Y(m)) := \min_{Y \in [x(k), I(x(k), Dr)]}\{C_{Dr}^2(Y) + Cgm(x(k), Y, \overline{Dr})\};$

    IF $C(x(k)) + Cgm(x(k), Y(m)) + C_{Dr}^2(Y(m))\ <\ J_x$ THEN

        // Update trajectory:

        $k := k+1;$

$x(k):=Y(m);$

$k:=k+1;$

$x(k):= x(k-1)+\delta_{D_r}(-d_1,-d_2);$

// Update upper bound on the optimal cost:

$J_X := C(x(k))+Cgm(x(k),Y(m))+C^2_{D_r}(Y(m));$

// Update trajectory cost up to the point $x(k)$:

$C(x(k)):= C(x(k))+Cgm(x(k-2),x(k-1))+Cgm(x(k-1),x(k));$

    ELSE

        // Cannot add another step to the current trajectory.

        // Then the trajectory $x(k)$ is optimal.

        STOP := TRUE;

    END

END DO

$swap(Dr,\overline{Dr});$

$k:=k+1;$

$x(k):=I(x(k),Dr);$

$k:=k+1;$

$x(k):= x(k-1)+\delta_{\overline{D_r}}(-d_1,-d_2);$

$k:=k+1;$

$x(k):= P_{D_r};$

//Optimal trajectory cost:

$C(x(k)):= J_X;$

To help understand the DSA algorithm, we illustrate it step by step using the following example. Throughout the example, the reader is referred to Figure 4.10. Given an initial surplus vector $x(0)$, we go through the following steps:

Step 1: Compute $C_1^2(x(0))$ and $C_2^2(x(0))$. Say that $C_2^2(x(0)) < C_1^2(x(0))$, and let J_x be the upper bound on the optimal trajectory cost and the initial direction of search is Direction 2. That is, $J_x = C_2^2(x(0))$ and Dr=2. Now, we know that the optimal trajectory is at least $D_2 - 2 - step$. Then, we try to add another step to the trajectory and see if we can lower the upper bound J_x. This we do in the next step.

Step 2: for all $Y \in [x(1), I(x(1),2)]$, find the best $D_1 - 2 - step$ trajectory that starts from Y and leads to the cyclic schedule at Point B. To do this, we perform a line search along Segment $[x(1), I(x(1),2)]$, using the golden section algorithm (see Bazaraa and Shetty (1979)). The optimal $D_1 - 2 - step$ trajectory starting at an initial point on Segment $[x(1), I(x(1),2)]$ is obtained for Point $Y(1)$ with cost $C_1^2(Y(1))$. Now, we check whether it is optimal to add this $D_1 - 2 - step$ trajectory to the trajectory up to Point $x(1)$ and obtain a $D_2 - 3 - step$ trajectory with a lower cost. The cost of this new $D_2 - 3 - step$ trajectory is $C_2^3(x(0)) = C(x(1)) + Cgm(x(1), Y(1)) + C_1^2(Y(1))$. Where $C(x(1))$ is the cost of the trajectory up to Point $x(1)$ starting at Point $x(0)$, and $Cgm(x(1), Y(1))$ is the cost of Segment $[x(1), Y(1)]$. Now, suppose that $C_2^3(x(0)) < J_x$. Then, it is better to move along this new $D_2 - 3 - step$ trajectory. The new upper bound on the optimal trajectory cost is $J_x = C_2^3(x(0))$. In the next step we check whether it is possible to add another step to the current trajectory and reach the cyclic schedule at Point D.

Step 3: Let $x(2) = Y(1)$ and $x(3) = x(2) + \delta_1(-d_1, -d_2)$. Now, the situation is identically the same as if we started with Point $x(1)$. Therefore, we need to go through Step 2 once again. Suppose, we did that and found a new upper bound $J_x = C_2^4(x(0))$. Then, it is an improvement to add a new step to the trajectory and obtain a new trajectory $D_2 - 4 - step$.
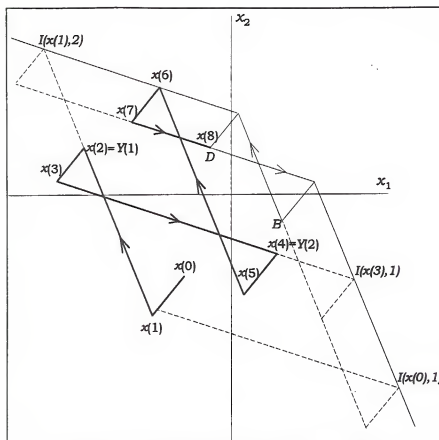
Figure 4.10: Illustration of the DSA algorithm.

$\underline{Step\ 4}$: Let $x(4)=Y(2)$ and $x(5)=x(3)+\delta_2(-d_1,-d_2)$. Similarly, we go through Step 2. Suppose, we did that and could not lower the upper bound on the optimal trajectory cost. Hence $D_2-4-step$ is optimal. To complete the trajectory, we let $x(6)=I(x(5),2)$, $x(7)=x(6)+\delta_1[-d_1,-d_2]$, and $x(8)=D$. The optimal trajectory is given by $\{x(0),x(1),...,x(8)\}$ and its cost is given by J_x.

Notice that each time we try to add a new step to the current trajectory, we sweep all possible trajectories in the direction of search and pick the one that minimizes the cost of reaching the cyclic schedule (this is done by means of the line search procedure). This is why the algorithm is called Direction Sweeping Algorithm.

In the following, we prove the validity of the line search and the uniqueness of its solution. For this, we use the algorithm notation.

Proposition 4.2: The cost function $C_{Dr}^2(x)$ is strictly convex in x. Where, $x \in [x(k), I(x(k), \overline{Dr})]$.

Proof: Let

$$C_{Dr}^2(x) = Cgm(x, x', 0) + Cgm(x', I(x', Dr), Dr) + Cgm(I(x', Dr), x'', 0) + Cgm(x'', P_{\overline{Dr}}, \overline{Dr}).$$

Where, $Cgm(X, Y, l)$ is the cost of Segment $[X, Y]$ when direction l is followed. Here, direction 0 is the South-West direction. In other words, the machine is being set up. In general, we can write the cost along Segment $[X, Y]$ in x-space as follows:

$$Cgm(X, Y, l) = \sum_{i=1}^{2} \frac{1}{2} \frac{h_i}{(\sigma_i U_i - d_i)} \{y_i^2 - x_i^2\}$$

where $X = (x_1, x_2)$, $Y = (y_1, y_2)$ and $h_i = c_i^+$, if x_i and y_i are positive and $h_i = -c_i^-$, if x_i and y_i are negative for $i=1,2$.

If Y is fixed, it is easy to show $Cgm(X, Y, l)$ is a strictly convex function of X. In particular, $Cgm(x'', P_{\overline{Dr}}, \overline{Dr})$ is a strictly convex function of x'' for fixed $P_{\overline{Dr}}$ (recall that P_{Dr} is either Point B or Point D on the cyclic schedule). Since x'' can be expressed as a linear function of $P_{\overline{Dr}}$, $Cgm(I(x', Dr), x'', 0) + Cgm(x'', P_{\overline{Dr}}, \overline{Dr})$ is a strictly convex function of $I(x', Dr)$ for fixed $P_{\overline{Dr}}$. Continuing the reasoning in the same manner, and since the domain $[x(k), I(x(k), \overline{Dr})]$ is convex, it follows that $C_{Dr}^2(x)$ is a strictly convex function of $x \in [x(k), I(x(k), \overline{Dr})]$. ∎

Using Proposition 4.2, it follows that the line search in the algorithm always provides a unique optimal solution.

Theorem 4.2: The DSA algorithm gives a unique optimal trajectory with the optimal cost $C_{i^*}^{m^*}(x(0))$, where $x(0)$ is the initial point of the trajectory in x-space, m^* is the optimal number of steps in the trajectory and i^* is the initial direction to follow.

Proof: To see that the trajectory obtained by the DSA is indeed optimal, consider an initial point $x(0)$. Then, the DSA algorithm can be described by the following recursive expressions:

$$C_{i^*}^2(x(0)) = \min_{i=1,2}\{C_i^2(x(0))\};$$

$$C_{i^*}^{m+1}(x(0)) = \min\left\{C_{i^*}^m(x(0)); \min_{Y \in [x(2m-3), J(x(2m-3), i_m)]}\{C(Y) + C_{\bar{i}_m}^2(Y)\}\right\}$$

where $m = 2, 3, \ldots, m^*$; $i_1 = \overline{i^*}$, $i_2 = i^*$, $i_3 = \overline{i^*}$, ...; i^* is the starting direction of search, $C_{i^*}^m(x(0))$ is the cost of the D_{i^*}-m-step trajectory and $C(Y)$ is the cost of the current trajectory up to point Y. The remaining quantities are the same as defined by the algorithm. These recurrence relation and Proposition 4.2 guarantee that each step added to the trajectory is optimal. Therefore, by the principle of optimality, the trajectory obtained is optimal. Furthermore, by Proposition 4.2 the optimal trajectory is unique.∎

Figure 4.11 shows an optimal trajectory starting at Point $(-10, -10)$ for the following problem data: $d_1 = 2$, $d_2 = 3$, $U_1 = 6$, $U_2 = 6$, $\delta_1 = 2$, $\delta_2 = 1$, $c_1^+ = 10$, $c_2^+ = 10$, $c_1^- = 50$, and $c_2^- = 50$. Notice that, the optimal trajectory is $D_1 - 5 - step$.

Figure 4.12 shows different optimal trajectories starting at different initial points. Notice that, the setup switching policy is a kind of corridor policy. The walls of the corridors have negative slopes with respect to the axis. Also, the corridor is linear. Which means that the boundaries in Region $\Re^u$ are linear as stated in Fact 4.2.
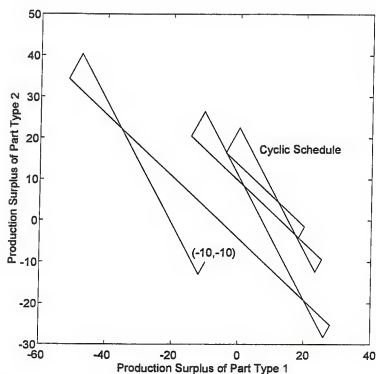
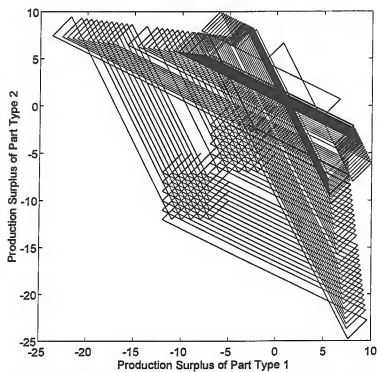Figure 4.11: Optimal trajectory obtained by the DSA algorithm.

Figure 4.12: Linearity of the boundaries in Region $\Re^u$.

4.6 Optimal Transient Solution in Region $\Re^u$ (Analytical Approach)

In this section, we derive the optimal transient solution in Region $\Re^u$ analytically, based on the DSA algorithm of the previous section. Together with the optimal solution of Region $\Re^o$, we obtain the complete analytical optimal transient solution. The algorithmic approach can be used when one tries to extend the problem to the multi-part-type case. The analytical approach is always desirable especially in the real-time scheduling context. It usually involves the use of a formula, which is an excellent tool for studying extreme cases where one is interested in the behavior of the solution with respect to one or many parameters of the problem. In addition, it provides insights to the more general case and indicates the pertinent parameters of the system.

4.6.1 Preliminary discussion and results

According to Fact 4.4, to reach the cyclic schedule in finite time, we need to perform a setup on one of the two boundaries L12 or L21. However, starting from a point in $\Re^u$, we may need more than one setup to bring the surplus trajectory to the cyclic schedule. For instance, as illustrated in Figure 4.13, starting at Point 0 in $\Re^u$ with an initial setup change for Part Type 2, we have a choice to change the setup at Point 1 on the boundary L21 and reach the cyclic schedule at Point D, or have setup at Point 2 and then at Point 3 on L12 and reach the cyclic schedule at Point B. If the cost can be reduced, we may want to initiate a setup on the segment [2',3] and swing back to boundary L21. As a consequence, the optimal trajectory leading to the cyclic schedule in finite time will be at least $D_i - 2 - step$, since we need a setup at the initial point and one more, once the boundary of Region $\Re^o$ is reached.
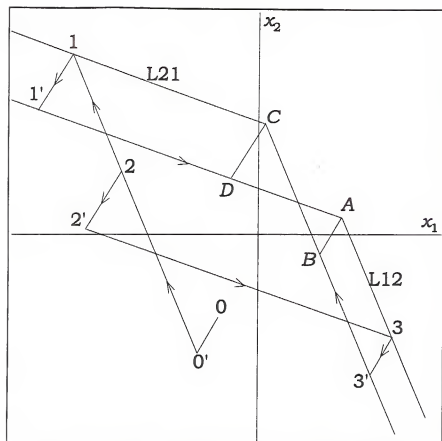
Figure 4.13: Illustration of the setup switching policy

Figure 4.12 of the previous section shows the application of the DSA algorithm to an example where optimal trajectories starting at different initial points are shown. Notice that the setup switching policy is a special corridor policy (see Sharifnia et al. (1991)) with two windows. Based on numerical results, we observe that the general optimal setup switching policy is as illustrated in Figure 4.14. In this case, the surplus trajectory will bounce back and forth between the two walls of the corridor ($\mathcal{B}1$ and $\mathcal{B}2$ in Figure 4.14) until the trajectory passes through one of the windows of the corridor. When this happens, the trajectory will encounter either Line L12 or Line L21 then bounces back and follows either Line L2 or Line L1 and reaches the cyclic schedule at Point D or Point B (Figure 4.14). Therefore, the optimal setup switching policy is completely characterized when the corridor walls $\mathcal{B}1$ and $\mathcal{B}2$

and the corridor windows defined by Points *W1*, *W2*, *V1* and *V2* in *x*-space are determined.

The purpose of this section is to obtain, analytically, the optimal setup switching policy in Region $\mathfrak{R}^u$. This is achieved by determining the equations of the boundaries $\mathcal{B}1$ and $\mathcal{B}2$ as well as Points *W1*, *W2*, *V1* and *V2*. One of the advantages of the analytical solution is that, once we have the equations of Boundaries $\mathcal{B}1$ and $\mathcal{B}2$, we eliminate the line search procedure in the DSA algorithm and substitute it with an intersection point calculation, which is clearly much faster and much simpler to implement (as a formula). Before we procedure any further, we state some preliminary results that are needed in the subsequent analyses.
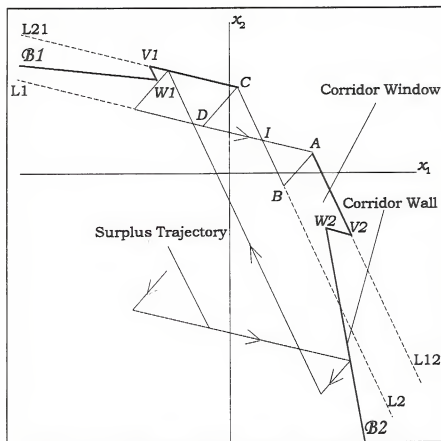

Figure 4.14: Illustration of the setup switching policy.

We denote by Line $L(Y)$, the line going through a point $Y = (y_1, y_2)$ in x-space and parallel to Line L2. We denote by Line $L'(Y)$, the line going through a point $Y = (y_1, y_2)$ in x-space and parallel to Line L1. The equations of $L(Y)$ and $L'(Y)$ are given by

$$\text{Line } L(Y): \ d_2(1 - \rho_2)(x_1 - y_1) + d_1\rho_2(x_2 - y_2) = 0; \tag{4.11}$$

$$\text{Line } L'(Y): \ d_2\rho_1(x_1 - y_1) + d_1(1 - \rho_1)(x_2 - y_2) = 0. \tag{4.12}$$

Let $Z_1 = (z_{11}, z_{12})$ be the intersection point of $L(Y)$ with L21 and $Z_2 = (z_{21}, z_{22})$ be the intersection point of $L'(Y)$ and L12. Then, Z_1 and Z_2 are given as follows:

$$z_{11} = \frac{(1 - \rho_1)}{(1 - \rho)}\left\{(1 - \rho_2)y_1 + (d_1/d_2)\rho_2 y_2 - \frac{\rho_1\rho_2 C_1}{(1 - \rho_1)} - (d_1/d_2)\rho_2 C_2\right\}; \tag{4.13a}$$

$$z_{12} = \frac{(1 - \rho_2)}{(1 - \rho)}\left\{-(d_2/d_1)\rho_1 y_1 - \frac{\rho_1\rho_2}{(1 - \rho_2)}y_2 + (d_2/d_1)\rho_1 C_1 + (1 - \rho_1)C_2\right\}; \tag{4.13b}$$

$$z_{21} = \frac{(1 - \rho_1)}{(1 - \rho)}\left\{-\frac{\rho_1\rho_2}{(1 - \rho_1)}y_1 - (d_1/d_2)\rho_2 y_2 + (1 - \rho_2)A_1 + (d_1/d_2)\rho_2 A_2\right\}; \tag{4.14a}$$

$$z_{22} = \frac{(1 - \rho_2)}{(1 - \rho)}\left\{(d_2/d_1)\rho_1 y_1 + (1 - \rho_1)y_2 - (d_2/d_1)\rho_1 A_1 - \frac{\rho_1\rho_2 A_2}{(1 - \rho_2)}\right\}. \tag{4.14b}$$

We denote by $C_{X,Y}$ the cost of Segment $[X, Y]$ in x-space, when X and Y are within the same quadrant in x-space. For $X = (x_1, x_2)$ and $Y = (y_1, y_2)$, we have

$$C_{X,Y} = \sum_{i=1}^{2}\frac{h_i}{2(\sigma_i U_i - d_i)}\left\{y_i^2 - x_i^2\right\} \tag{4.15}$$

where $h_i = c_i^+$ if x_i and y_i are positive and $h_i = -c_i^-$ if x_i and y_i are negative for $i = 1, 2$.

Remark: If X and Y are not in the same quadrant, we can always break the segment into two or three parts each contained within one quadrant in x-

space. Then, the total $[X, Y]$ segment cost is the sum of the two or three segment costs.

4.6.2 Equations of Boundaries $\mathcal{B}1$ and $\mathcal{B}2$

In this subsection, we determine the locus of the points where it might be optimal to initiate a setup for the other part type. In other words, how far should we go with the production of the current part type before we switch to the other. Notice that, at this point, we don't know whether this switch will lower the cost of the current trajectory or not. This will be decided later on, when we determine Points $W1$, $W2$, $V1$ and $V2$. Indeed, if the switch happens to be within the corridor windows, then it is optimal not to switch and continue producing the current part type until either Boundary L12 or Boundary L21 is encountered, where it is optimal to switch to the other part type and reach the cyclic schedule at Point B or Point D.

To determine the equations representing the corridor walls in x-space, we start with an initial surplus point in the third quadrant of the x-space. It will be shown that the equations of the corridor walls (Boundaries $\mathcal{B}1$ and $\mathcal{B}2$) do not depend on the initial surplus. To be able to follow the derivation, the reader is referred to Figure 4.15. First, we derive the equation of Boundary $\mathcal{B}1$. Before we start the derivation, we need to determine the coordinates of Points $X = (x_1, x_2)$ and $V = (v_1, v_2)$, given $X_0 = (x_1^0 + \delta_2 d_1, x_2^0 + \delta_2 d_2)$ (without loss of generality). In this case, X belongs to Line L(X_0') and V is the intersection of Line L12 and Line L'(X), where $X_0' = (x_1^0, x_2^0)$. Then, using (4.11), (4.14a) and (4.14b), we have:
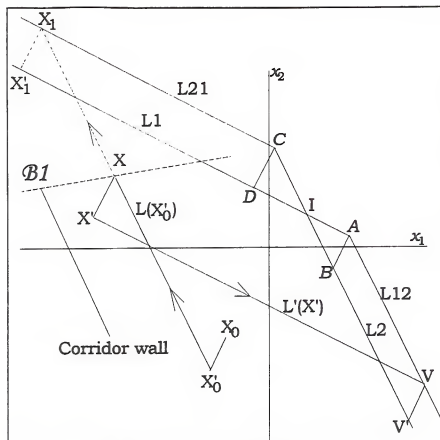
Figure 4.15: Corridor walls

$$x_2 = (d_2/d_1)\frac{(1-\rho_2)}{\rho_2}(x_1 - x_1^0) + x_2^0;$$

$$v_1 = \frac{(1-\rho_1)}{(1-\rho)}\left\{-\frac{\rho_1\rho_2}{(1-\rho_1)}(x_1 - \delta_1 d_1) - (d_1/d_2)\rho_2(x_2 - \delta_1 d_2) + (1-\rho_2)A_1 + (d_1/d_2)\rho_2 A_2\right\};$$

$$v_2 = \frac{(1-\rho_2)}{(1-\rho)}\left\{(d_2/d_1)\rho_1(x_1 - \delta_1 d_1) + (1-\rho_1)(x_2 - \delta_1 d_2) - (d_2/d_1)\rho_1 A_1 - \frac{\rho_1\rho_2 A_2}{(1-\rho_2)}\right\}.$$

As mentioned before, the objective is to determine the location in x-space of the switching point X that minimizes the cost of the trajectory emanating at Point X_0 in x-space and reaching the cyclic schedule at Point B (Figure 4.15). Notice that the current trajectory in Figure 4.15 is a $D_2 - 2 - step$ trajectory given by $\{X_0, X_0', X_1, X_1', D, I\}$. Now, if we introduce a setup switch at Point X, the new trajectory will be a $D_2 - 3 - step$ trajectory given by $\{X_0, X_0', X, X', V, V', B, I\}$.

At this stage, we have no guarantee that the new trajectory will have a lower cost than the current one. This will be decided later when we determine *W1*, *W2*, *V1* and *V2*. All we are saying at this point, is that if switching to the other part type would lower the cost of reaching the cyclic schedule, then the optimal switch must happen at Point *X* in *x*-space.

Using the notation of Proposition 4.1, $C_2^3(X_0)$ would be the cost of Trajectory $\{X_0, X_0', X, X', V, V', B, I\}$. Using (4.15), we have

$$C_2^3(X_0) = \frac{1}{2d_1(1-\rho_1)}\left\{(x_1 - \delta_1 d_1)^2 c_1^- + v_1^2 c_1^+\right\} + \frac{1}{2d_2(1-\rho_2)}\left\{x_2^2 c_2^+ + (v_2 - \delta_2 d_2)^2 c_2^-\right\}$$

$$- \frac{1}{2d_1}\left\{(x_1^0 + \delta_2 d_1)^2 c_1^- + I_1^2 c_1^+\right\} - \frac{(x_2^0)^2 c_2^-}{2d_2(1-\rho_2)} - \frac{1}{2d_2}(x_2^0 + \delta_2 d_2)^2 c_2^- + \frac{\rho_2}{2d_2(1-\rho_2)}I_2^2 c_2^+.$$

Notice that, if we substitute x_2, v_1, v_2 by their expressions as a function of x_1, $C_2^3(X_0)$ will be a quadratic function of x_1. Hence, $C_2^3(X_0)$ may have a unique minimum. A necessary condition for x_1 to be a minimum of $C_2^3(X_0)$ (and sufficient in this case) is given by

$$\frac{dC_2^3(X_0)}{dx_1} = 0.$$

Differentiating x_2, v_1, v_2 with respect to x_1, we get

$$\frac{dx_2}{dx_1} = -(d_2/d_1)\frac{(1-\rho_2)}{\rho_2};$$

$$\frac{dv_1}{dx_1} = 1;$$

$$\frac{dv_2}{dx_1} = -(d_2/d_1)\frac{(1-\rho_2)}{\rho_2}.$$

Now, differentiating $C_2^3(X_0)$ with respect to x_1 and expressing v_1 and v_2 as a function of x_1 and x_2 and setting the result to 0, gives the following equation of a line in *x*-space:

$$\mathcal{B}1: \ M_1 x_1 + M_2 x_2 + M_3 = 0 \tag{4.16}$$

where

$$M_1 = \frac{1}{(1-\rho_1)} c_1^- - \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho)} c_1^+ - (d_1/d_2)(\rho_1/\rho_2) \frac{(1-\rho_2)}{(1-\rho)} c_2^-; \tag{4.17a}$$

$$M_2 = -(d_1/d_2) \frac{\rho_2}{(1-\rho)} c_1^+ - \frac{1}{\rho_2} c_2^+ - \frac{(1-\rho_1)(1-\rho_2)}{\rho_2 (1-\rho)} c_2^-; \tag{4.17b}$$

$$M_3 = -(M_1 C_1 + M_2 C_2). \tag{4.17c}$$

Notice that $\mathcal{B}1$ contains Point C on the cyclic schedule. This result is not a surprise since, if we started with an initial surplus X_0 belonging to Line L12, Point C on the cyclic schedule would realize the minimum for $C_2^3(X_0)$ and therefore is a candidate for initiating a setup.

With a similar procedure, we obtain the equation of Boundary $\mathcal{B}2$, where it is possible to switch from Part Type 1 to Part Type 2.

$$\mathcal{B}2: \ N_1 x_1 + N_2 x_2 + N_3 = 0 \tag{4.18}$$

where

$$N_1 = -(d_2/d_1) \frac{\rho_1}{(1-\rho)} c_2^- - \frac{1}{\rho_1} c_1^+ - \frac{(1-\rho_1)(1-\rho_2)}{\rho_1 (1-\rho)} c_1^-; \tag{4.19a}$$

$$N_2 = \frac{1}{(1-\rho_2)} c_2^- - \frac{\rho_1 \rho_2}{(1-\rho_2)(1-\rho)} c_2^+ - (d_2/d_1)(\rho_2/\rho_1) \frac{(1-\rho_1)}{(1-\rho)} c_1^-; \tag{4.19b}$$

$$N_3 = -(N_1 A_1 + N_2 A_2). \tag{4.19c}$$

It is worth mentioning that in the previous section, we observed from numerical solutions that the optimal setup switching boundaries in x-space are linear. It is verified here by the equations of the corridor walls.

4.6.3 Corridor windows

In this section, we determine the corridor windows, which are characterized by the coordinates of Points $W1 = (w_{11}, w_{12})$, $W2 = (w_{21}, w_{22})$, $V1 = (v_{11}, v_{12})$ and $V2 = (v_{21}, v_{22})$ in x-space (Figure 4.14). $W1$ and $W2$ are the points on the corridor boundaries where an additional setup does not reduce the cost of reaching the cyclic schedule. Geometrically, this means that beyond these two points $W1$ and $W2$ on the boundaries $\mathcal{B}1$ and $\mathcal{B}2$, it would not be optimal to initiate a setup (that is, the cost of the current trajectory cannot be lowered). First, we derive the coordinates of $W1$ and $W2$. The coordinates of $V1$ and $V2$ are obtained as a function of those of $W1$ and $W2$.

We derive the coordinates of $W1$ first. To be able to follow the derivation the reader is referred to Figure 4.16. As we mentioned above, $W1$ is the point in x-space that realizes the equilibrium between the cost of Trajectory $\{W1, V1, V1', I\}$ and the cost of Trajectory $\{W1, W1', Z1, Z1', I\}$. $W1$ must be a point of $\mathcal{B}1$ since we know that $\mathcal{B}1$ is the set of optimal locations where we switch to Part Type 2. Figure 4.16 shows an example of three trajectories τ_1, τ_2 and τ_3. For trajectory τ_1, it is optimal to switch to Part Type 1. For trajectory τ_2, we are indifferent between switching to Part Type 1 or continuing the production of Part Type 2, switching setup at Point $V1$ and then reaching the cyclic schedule at Point D. In this case the two trajectories have the same cost. For trajectory τ_3, it is not optimal to switch to Part Type 1 and it is optimal to continue the production of Part Type 2 and reach the cyclic schedule at Point D.

Before we start the derivation, we need the coordinates of Point $Z1 = (z_{11}, z_{12})$ shown in Figure 4.16. In this case, $V1$ belongs to $L(W1) \cap L21$ and $Z1$ belongs to $L'(W1') \cap L12$. Once we have $V1$ and $Z1$, $V1'$ and $Z1'$ are easily obtained. $V1$ and $Z1$ are defined as follows:
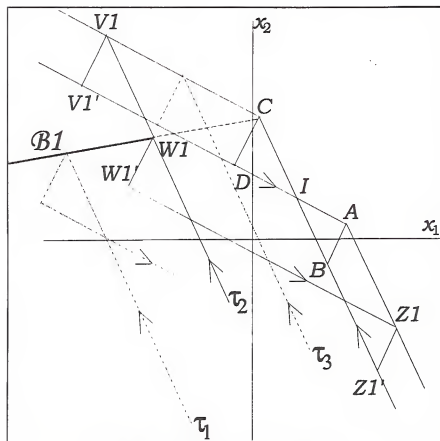
Figure 4.16: Corridor windows

$$z_{11} = \eta_{11}w_{11} + v_{11}; \tag{4.20a}$$

$$z_{12} = \eta_{12}w_{12} + v_{12}; \tag{4.20b}$$

$$v_{11} = \gamma_{11}w_{11} + \mu_{11}; \tag{4.21a}$$

$$v_{12} = \gamma_{12}w_{12} + \mu_{12} \tag{4.21b}$$

where

$$\eta_{11} = \frac{\rho_2}{(1-\rho)}\left\{(d_1/d_2)(M_1/M_2)(1-\rho_1) - \rho_1\right\}; \tag{4.22a}$$

$$\eta_{12} = \frac{(1-\rho_2)}{(1-\rho)}\left\{(d_2/d_1)\rho_1 - (M_1/M_2)(1-\rho_1)\right\}; \tag{4.22b}$$

$$v_{11} = \frac{(1-\rho_1)(1-\rho_2)}{(1-\rho)}A_1 + \frac{(1-\rho_1)\rho_2}{(1-\rho)}(d_1/d_2)(A_2 + (M_3/M_2)) + \frac{\rho_2}{(1-\rho)}\delta_1 d_1; \tag{4.23a}$$

$$V_{12} = \frac{(1-\rho_2)}{(1-\rho)}\left\{-\rho_1(d_2/d_1)A_1 - \frac{\rho_1\rho_2}{(1-\rho_2)}A_2 - (1-\rho_1)(M_3/M_2) - (d_2/d_1)\delta_1 d_1\right\}; \quad (4.23b)$$

$$\gamma_{11} = \frac{(1-\rho_1)}{(1-\rho)}\left\{(1-\rho_2) - \rho_2(d_1/d_2)(M_1/M_2)\right\}; \quad (4.24a)$$

$$\gamma_{12} = \frac{\rho_1}{(1-\rho)}\left\{\rho_2(M_1/M_2) - (1-\rho_2)(d_2/d_1)\right\}; \quad (4.24b)$$

$$\mu_{11} = -\frac{\rho_1\rho_2}{(1-\rho)}C_1 - \frac{\rho_2(1-\rho_1)}{(1-\rho)}(d_1/d_2)(C_2 + (M_3/M_2)); \quad (4.25a)$$

$$\mu_{12} = \frac{\rho_1(1-\rho_2)}{(1-\rho)}(d_2/d_1)C_1 + \frac{(1-\rho_1)(1-\rho_2)}{(1-\rho)}C_2 + \frac{\rho_1\rho_2}{(1-\rho)}(M_3/M_2). \quad (4.25b)$$

Denote by TC_{11}, the cost of Trajectory $\{W1, V1, V1', I\}$ and by TC_{12} the cost of Trajectory $\{W1, W1', Z1, Z1', I\}$. Using (4.15), we have

$$TC_{11} = \frac{1}{2}\left\{-\frac{w_{11}^2 c_1^-}{d_1} - \frac{I_2^2 c_2^*}{d_2} + \frac{\left[(v_{11} - \delta_1 d_1)^2 c_1^- + \rho_1 I_1^2 c_1^*\right]}{d_1(1-\rho_1)} + \frac{\left[v_{12}^2 c_2^* - \rho_2 w_{12}^2 c_2^*\right]}{d_2(1-\rho_2)}\right\}; \quad (4.26)$$

$$TC_{12} = \frac{1}{2}\left\{-\frac{\left[w_{11}^2 c_1^- + I_1^2 c_1^*\right]}{d_1} + \frac{\left[(w_{11} - \delta_1 d_1)^2 c_1^- + z_{11}^2 c_1^*\right]}{d_1(1-\rho_1)}\right.$$

$$\left. + \frac{w_{12}^2 c_2^*}{d_2} + \frac{\left[(z_{12} - \delta_2 d_2)^2 c_2^- + \rho_2 I_2^2 c_2^*\right]}{d_2(1-\rho_2)}\right\}. \quad (4.27)$$

As we mentioned above, $W1 = (w_{11}, w_{12})$ is the point in x-space for which we have $TC_{11} = TC_{12}$. Furthermore, $W1$ belongs to the boundary $\mathcal{B}1$. Therefore, $W1$ should satisfy the following two conditions:

$$\begin{cases} TC_{11}(w_{11}) - TC_{12}(w_{11}) = 0; \\ w_{12} = -(M_1/M_2)w_{11} - (M_3/M_2). \end{cases} \quad (4.28)$$

Notice that, substituting v_{11}, v_{12}, z_{11}, z_{12} and w_{12} by their expressions as a function of w_{11}, makes TC_{11} and TC_{12} depend on w_{11} only. After manipulation, we obtain the following system of equations:

$$\begin{cases} \alpha_1 w_{11}^2 + \alpha_2 w_{11} + \alpha_3 = 0; \\ w_{12} = -(M_1/M_2)w_{11} - (M_3/M_2) \end{cases} \qquad (4.29)$$

where

$$\alpha_1 = \frac{(\gamma_{11}^2 - 1)c_1^- - \eta_{11}^2 c_1^+}{2d_1(1 - \rho_1)} + \frac{(\gamma_{12}^2 - (M_1/M_2)^2)c_2^* - \eta_{12}^2 c_2^-}{2d_2(1 - \rho_2)}; \qquad (4.30a)$$

$$\alpha_2 = \frac{(\gamma_{11}(\mu_{11} - \delta_1 d_1) + \delta_1 d_1)c_1^- - \eta_{11} v_{11} c_1^+}{d_1(1 - \rho_1)}$$

$$+ \frac{\left(\gamma_{12}\mu_{12} - \dfrac{M_1 M_3}{M_2^2}\right)c_2^* - \eta_{12}(v_{12} - \delta_2 d_2)c_2^-}{d_2(1 - \rho_2)}; \qquad (4.30b)$$

$$\alpha_3 = \frac{\mu_{11}(\mu_{11} - 2\delta_1 d_1)c_1^- - (v_{11}^2 - I_1^2)c_1^+}{2d_1(1 - \rho_1)}$$

$$+ \frac{(\mu_{12}^2 - (M_3/M_2)^2 - I_2^2)c_2^* - (v_{12} - \delta_2 d_2)^2 c_2^-}{2d_2(1 - \rho_2)}. \qquad (4.30c)$$

For most practical situations (finite c_i^+ and c_i^- ($i = 1, 2$), it can be shown that $W1$ is located in the second quadrant of the x-space. Hence, w_{11} in (4.29) must be the negative root.

In a similar fashion, we determine the coordinates of the Point $W2$ in x-space, by solving the following system of two equations:

$$\begin{cases} \beta_1 w_{22}^2 + \beta_2 w_{22} + \beta_3 = 0; \\ w_{21} = -(N_2/N_1)w_{22} - (N_3/N_1) \end{cases} \qquad (4.31)$$

where

$$\beta_1 = \frac{(\gamma_{21}^2 - (N_2/N_1)^2)c_1^+ - \eta_{21}^2 c_1^-}{2d_1(1-\rho_1)} + \frac{(\gamma_{22}^2 - 1)c_2^- - \eta_{22}^2 c_2^+}{2d_2(1-\rho_2)}; \tag{4.32a}$$

$$\beta_2 = \frac{\left(\gamma_{21}\mu_{21} - \frac{N_2 N_3}{N_1^2}\right)c_1^+ - \eta_{21}(\nu_{21} - \delta_1 d_1)c_1^-}{d_1(1-\rho_1)}$$

$$+ \frac{(\gamma_{22}(\mu_{22} - \delta_2 d_2) + \delta_2 d_2)c_2^- - \eta_{22}\nu_{22}c_2^+}{d_2(1-\rho_2)}; \tag{4.32b}$$

$$\beta_3 = \frac{\left(\mu_{21}^2 - (N_3/N_1) - I_1^2\right)c_1^+ - (\nu_{21} - \delta_1 d_1)^2 c_1^-}{2d_1(1-\rho_1)}$$

$$+ \frac{\mu_{22}(\mu_{22} - 2\delta_2 d_2)c_2^- - (\nu_{22}^2 - I_2^2)c_2^+}{2d_2(1-\rho_2)}; \tag{4.32c}$$

$$\eta_{21} = \frac{(1-\rho_1)}{(1-\rho)}\{(d_1/d_2)\rho_2 - (N_2/N_1)(1-\rho_2)\}; \tag{4.33a}$$

$$\eta_{22} = \frac{\rho_1}{(1-\rho)}\{(d_2/d_1)(N_2/N_1)(1-\rho_2) - \rho_2\}; \tag{4.33b}$$

$$\nu_{21} = \frac{(1-\rho_1)}{(1-\rho)}\left\{-\frac{\rho_1\rho_2}{(1-\rho_1)}C_1 - \rho_2(d_1/d_2)C_2 - (1-\rho_2)(N_3/N_1) - (d_1/d_2)\delta_2 d_2\right\}; \tag{4.34a}$$

$$\nu_{22} = \frac{(1-\rho_2)\rho_1}{(1-\rho)}(d_2/d_1)(C_1 + (N_3/N_1)) + \frac{(1-\rho_1)(1-\rho_2)}{(1-\rho)}C_2 + \frac{\rho_1}{(1-\rho)}\delta_2 d_2; \tag{4.34b}$$

$$\gamma_{21} = \frac{\rho_2}{(1-\rho)}\{\rho_1(N_2/N_1) - (1-\rho_1)(d_1/d_2)\}; \tag{4.35a}$$

$$\gamma_{22} = \frac{(1-\rho_2)}{(1-\rho)}\{(1-\rho_1) - \rho_1(d_2/d_1)(N_2/N_1)\}; \tag{4.35b}$$

$$\mu_{21} = \frac{(1-\rho_1)(1-\rho_2)}{(1-\rho)}A_1 + \frac{\rho_2(1-\rho_1)}{(1-\rho)}(d_1/d_2)A_2 + \frac{\rho_1\rho_2}{(1-\rho)}(N_3/N_1); \tag{4.36a}$$

$$\mu_{22} = -\frac{\rho_1(1-\rho_2)}{(1-\rho)}(d_2/d_1)(A_1 + (N_3/N_1)) - \frac{\rho_1\rho_2}{(1-\rho)}A_2. \tag{4.36b}$$

w_{22} in (4.31) must be the negative root.

4.6.4 The complete analytical solution

At this stage, we have the complete analytical solution of the optimal production and setup control policy in Region $\mathfrak{R}^u$.

For initial surplus $x = (a,b)$ in Region $\mathfrak{R}^u$, the following procedure gives the optimal trajectory of the surplus levels in x-space.

Procedure 1

STEP 0:

    $X_1(0):= a$, $X_2(0):= b$;

    $k:=1$;

    IF $C_1^0(X(0)) \geq C_2^0(X(0))$ *THEN*

        $X_1(k):= X_1(0) - \delta_2 d_1$;

        $X_2(k):= X_2(0) - \delta_2 d_2$;

        GOTO STEP 1;

    ELSE

        $X_1(k):= X_1(0) - \delta_1 d_1$;

        $X_2(k):= X_2(0) - \delta_1 d_2$;

        GOTO STEP 2;

    ENDIF;

STEP 1:

    $k:=k+1$;

    $X_1(k):= \left(d_1\rho_2(M_3 + X_2(k-1)M_2) + d_2(1-\rho_2)X_1(k-1)M_2\right) \big/ \left(d_2(1-\rho_2)M_2 - d_1\rho_2 M_1\right)$;

    $X_2(k):= -(M_1/M_2)X_1(k) - (M_3/M_2)$;

    IF $X_1(k) \geq w_{11}$ *THEN*

        $X_1(k):= \gamma_{11}X_1(k-1) + \mu_{11}$;

        $X_2(k):= \gamma_{12}X_2(k-1) + \mu_{12}$;

        $k:=k+1$;

$$X_1(k) := X_1(k-1) - \delta_1 d_1;$$

$$X_2(k) := X_2(k-1) - \delta_1 d_2;$$

$$k := k+1;$$

$$X(k) := I;$$

GOTO STEP 3;

ELSE

GOTO STEP 2;

ENDIF;

STEP 2:

$$k := k+1;$$

$$X_2(k) := (d_2 \rho_1 (N_3 + X_1(k-1)N_1) + d_1(1-\rho_1)X_1(k-1)N_1)/(d_1(1-\rho_1)N_1 - d_2 \rho_1 N_2);$$

$$X_1(k) := -(N_2/N_1)X_2(k) - (N_3/N_1);$$

IF $X_2(k) \geq w_{22}$ *THEN*

$$X_1(k) := \gamma_{21}X_1(k-1) + \mu_{21};$$

$$X_2(k) := \gamma_{22}X_2(k-1) + \mu_{22};$$

$$k := k+1;$$

$$X_1(k) := X_1(k-1) - \delta_2 d_1;$$

$$X_2(k) := X_2(k-1) - \delta_2 d_2;$$

$$k := k+1;$$

$$X(k) := I;$$

GOTO STEP 3;

ELSE

GOTO STEP 1;

ENDIF;

STEP 3:

OUTPUT: OPTIMAL TRAJECTORY $= \{X(0),\ X(1),\ \cdots,\ X(k-1),\ X(k)\};$

Given the optimal trajectory of the surplus levels in x-space, it is not difficult to translate it into control actions.

In the following, we give two examples which are solved using Procedure 1. Example 1 illustrates a nonsymmetric case where product 2 is more expensive than product 1. For this example, we start with a backlog of 10 units of product 1 and with a zero inventory level of product 2. Example 2 illustrates a symmetric case where both products have the same importance. For this case, we start with a zero inventory level for both products. In each case, the optimal trajectory is given as a collection of points in x-space (points where the trajectory changes direction). The points shown in italic are the points at which the trajectory follows the cyclic schedule thereafter.

<u>Example 1</u>

Input:

$c_1^+ = 0.75$, $c_1^- = 7.5$, $U_1 = 7$, $d_1 = 3$, $\delta_1 = 1$;

$c_2^+ = 1.25$, $c_2^- = 18.75$, $U_2 = 5$, $d_2 = 2$, $\delta_2 = 1.5$.

Initial surplus $X(0)=(-10,0)$.

Output:

Optimal Trajectory ={ (-10,0), (-14.5,-3.0), (-52.8,35.3), (-55.8,33.3) , (24.0,-6.6), (19.5,-9.6), (-13.9,23.7), (-17.8,21.7), *(-2.3,14.4)=D,(22.7,1.9)=A,(18.2,-1.1)=B, (0.7,16.4)=C, (-2.3,14.4)=D, ...}.*

Initial setup change to Part Type 2.

<u>Example 2</u>

Input:

$c_1^+ = 1.0$, $c_1^- = 30.0$, $U_1 = 6$, $d_1 = 2$, $\delta_1 = 1$;

$c_2^+ = 1.0$, $c_2^- = 30.0$, $U_2 = 6$, $d_2 = 2$, $\delta_2 = 1$.

Initial surplus $X(0)=(0,0)$.

Output:

Optimal Trajectory $=\{$ (0,0), (-2.0,-2.0), (16.3,-14.2), (14.3,-16.2),

(-5.7,13.8), (-7.7,15.8), (12.6,-1.8), (10.6,-3.8), (-0.6,13.11),

(-2.6,11.1), *(-0.4,9.6)=D, (11.6,1.6)=A, (9.6,-0.4)=B, (1.6,11.6)=C,*

(-0.4,9.6)=D, ...}.

Initial setup change to Part Type 1.

In the following section, we study some extreme cases which are of either practical or theoretical importance. Since we have the analytical solution, it is relatively easy to study such cases.

4.6.5 Special cases

In this section, four special cases are investigated. Case 1: No backlog is allowed for Part Type 2 (without loss of generality). Case 2: No backlog is allowed for both parts. Case 3: No inventory is allowed for part type 2. Case 4: No inventory is allowed for both part types.

4.6.5.1 No backlog for Part Type 2

This case can be encountered in practice where a backlog for a Part Type is undesirable either because of customer requirements or because of a very high penalty incurred when the products are not delivered in time. To obtain the optimal policy in this case, we simply let the backlog instantaneous cost of Part Type 2 go to infinity. That is $c_2^- \to \infty$.

As before, we determine the corridor wall first, then we determine the windows of the corridors (if they exist).

<u>Corridor wall $\mathcal{B}1$</u>: dividing the equation of $\mathcal{B}1$ by c_2^- in (4.16) and taking the limit when $c_2^- \to \infty$, we get

$$\lim_{c_2^- \to \infty} M_1/c_2^- = -(d_1/d_2)(\rho_1/\rho_2)(1-\rho_2)/(1-\rho);$$

$$\lim_{c_2^- \to \infty} M_2/c_2^- = -(1-\rho_1)(1-\rho_2)/\rho_2(1-\rho);$$

$$\lim_{c_2^- \to \infty} M_3/c_2^- = (d_1/d_2)(\rho_1/\rho_2)((1-\rho_2)/(1-\rho))C_1 + ((1-\rho_1)(1-\rho_2)/\rho_2(1-\rho))C_2.$$

Rearranging terms in the expression of $\mathcal{B}1$, gives the following line equation:

$$d_2\rho_1(x_1 - C_1) + d_1(1-\rho_1)(x_2 - C_2) = 0.$$

Which is exactly the equation of Line L21. Hence, when backlog is not allowed for Part Type 2, Boundary $\mathcal{B}1$ becomes Line L21.

Corridor wall $\mathcal{B}2$: dividing the equation of $\mathcal{B}2$ by c_2^- in (4.18) and taking the limit when $c_2^- \to \infty$, we get

$$\lim_{c_2^- \to \infty} N_1/c_2^- = 0;$$

$$\lim_{c_2^- \to \infty} N_2/c_2^- = 1/(1-\rho_2);$$

$$\lim_{c_2^- \to \infty} N_3/c_2^- = -\delta_2 d_2/(1-\rho_2),$$

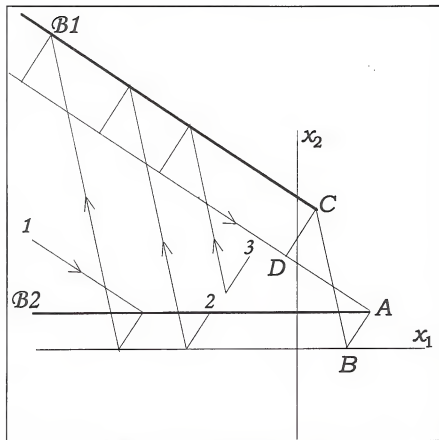which gives the following line equation for $\mathcal{B}2$:

$$x_2 = \delta_2 d_2.$$

This result was intuitively expected. Indeed, x_2 cannot be negative. Hence, the switch to Part Type 2 must occur exactly when its surplus reaches Level $\delta_2 d_2$, so that after the setup is completed, the surplus level of Part Type 2 is not negative.

Corridor Windows: To determine the corridor window, we divide expression (4.30a), (4.30b) and (4.30c) by c_2^- then take the limit when $c_2^- \to \infty$ and get:

$$\lim_{c_2^- \to \infty} \alpha_1/c_2^- = \alpha_1^\infty = -\left((M_1/M_2)^2 c_2^+ + \eta_{12}^2\right)/2d_2(1-\rho_2);$$

$$\lim_{c_2^- \to \infty} \alpha_2/c_2^- = \alpha_2^\infty = -\left((M_1/M_2)(M_3/M_2)c_2^+ + \eta_{12}(v_{12} - \delta_2 d_2)\right)/d_2(1-\rho_2);$$

Figure 4.17: Case where $c_2^- \to \infty$.

$$\lim_{c_2^- \to \infty} \alpha_3/c_2^- = \alpha_3^\infty = -\left((M_3/M_2)^2 c_2^+ + (\nu_{12} - \delta_2 d_2)^2\right)\big/2d_2(1 - \rho_2).$$

Now, if we compute the discriminant in the quadratic $\alpha_1^\infty w_{11}^2 + \alpha_2^\infty w_{11} + \alpha_3^\infty = 0$,

we get

$$(\alpha_2^\infty)^2 - 4(\alpha_1^\infty)(\alpha_3^\infty) = -4c_2^+\left[(M_1/M_2)(\nu_{12} - \delta_2 d_2) + (M_3/M_2)\eta_{12}\right]^2,$$

which is obviously negative. This means that the cost of reaching the cyclic schedule from Point D, is always less than that of reaching the limit cycle from Point B. Hence, the corridor window is the whole corridor wall given by the equation of Line L21. Figure 4.17 illustrates this policy.

4.6.5.2 No backlog for either part type

In this case, we have $c_1^- \to \infty$ and $c_2^- \to \infty$. In a similar fashion as the previous case, we can show that, $\mathcal{B}1$: $x_1 = \delta_1 d_1$ and $\mathcal{B}2$: $x_2 = \delta_2 d_2$. This means

that Region $\mathfrak{R}^u$ becomes "forbidden". That is, all initial surplus levels must be in Region $\mathfrak{R}^o$ (otherwise the cost would be infinite) and therefore, the optimal control actions of Region $\mathfrak{R}^o$ apply (see Figure 4.18).
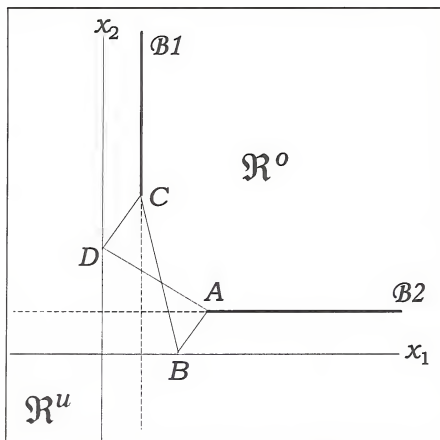

Figure 4.18: Case where $c_1^- \to \infty$ and $c_2^- \to \infty$.

<u>4.6.5.3 No inventory for Part Type 2</u>

This case happens in practice when there is no space available for a product and the latter has to be delivered to the customer at once after being produced. Another situation is when the inventory cost for the product is much higher than its backlog cost. In this situation it is better to deliver the product late than store it for a while before being delivered to the customer. This is usually known as "produce to order system" in the literature. To obtain the

optimal policy in this case, we simply let the inventory instantaneous cost of Part Type 2 go to infinity. That is $c_2^* \to \infty$.

Corridor wall $\mathcal{B}1$: dividing the equation of $\mathcal{B}1$ by c_2^* in (4.16) and taking the limit when $c_2^* \to \infty$, we get

$$\lim_{c_2^* \to \infty} M_1 / c_2^* = 0;$$

$$\lim_{c_2^* \to \infty} M_2 / c_2^* = -1/\rho_2;$$

$$\lim_{c_2^* \to \infty} M_3 / c_2^* = 0,$$

which gives $x_2 = 0$. This result was intuitively expected. Indeed, x_2 cannot be positive, hence the switch to Part Type 2 must occur when the backlog of that part type is completely eliminated.

Corridor wall $\mathcal{B}2$: dividing the equation of $\mathcal{B}2$ by c_2^* in (4.18) and taking the limit when $c_2^* \to \infty$, gives

$$\lim_{c_2^* \to \infty} N_1 / c_2^* = -(d_2/d_1)\rho_1/(1-\rho);$$

$$\lim_{c_2^* \to \infty} N_2 / c_2^* = -\rho_1\rho_2/(1-\rho_2)(1-\rho);$$

$$\lim_{c_2^* \to \infty} N_3 / c_2^* = (d_2/d_1)(\rho_1/(1-\rho))A_1 + (\rho_1\rho_2/(1-\rho_2)(1-\rho))A_2.$$

Rearranging terms in the expression of $\mathcal{B}2$, gives the following line equation:

$$d_2(1-\rho_2)(x_1 - A_1) + d_1\rho_2(x_2 - A_2) = 0,$$

which is exactly the equation of Line L12. Hence, when inventory is not allowed for Part Type 2, Boundary $\mathcal{B}2$ becomes Line L12.

Corridor Windows: To determine the corridor window, we divide expressions (4.32a), (4.32b) and (4.32c) by c_2^* then take the limit when $c_2^* \to \infty$ and get

$$\lim_{c_2^* \to \infty} \beta_1 / c_2^* = \beta_1^\infty = 0;$$

$$\lim_{c_2^* \to \infty} \beta_2/c_2^* = \beta_2^\infty = 0;$$

$$\lim_{c_2^* \to \infty} \beta_3/c_2^* = \beta_3^\infty = \left((1-\rho_2)/2(1-\rho)^2\right)d_2\delta_2^a,$$

which implies that $\beta_1^\infty w_{22}^2 + \beta_2^\infty w_{22} + \beta_3^\infty = \beta_3^\infty > 0$. Hence, the cost of reaching the cyclic schedule from Point B is always less than that of reaching the cyclic schedule from Point D. Therefore, the corridor window is the whole corridor wall given by the equation of Line L12. Figure 4.19 illustrates this policy.
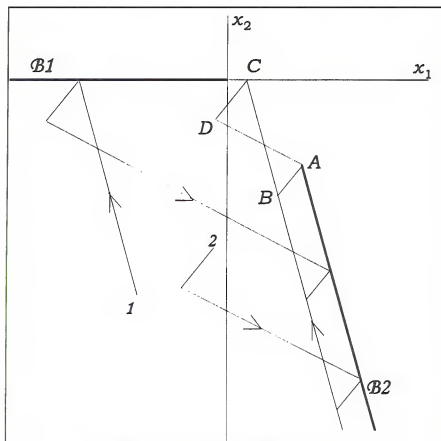

Figure 4.19: Case where $c_2^* \to \infty$.

4.6.5.4 No inventory for either part type

In this case, we have $c_1^* \to \infty$ and $c_2^* \to \infty$. In a similar fashion as in the previous case, we can show that, $\mathcal{B}1$: $x_1 = 0$ and $\mathcal{B}2$: $x_2 = 0$. This means that all

activities take place in the third quadrant of the x-space. Also, it is not difficult to show that when $c_1^* \to \infty$ and $c_2^* \to \infty$,

$$W1 \to C^\infty = \left(\delta_1 d_1 - \frac{(1-\rho_1)}{(1-\rho)} \delta d_1, 0 \right) \text{ and } W2 \to A^\infty = \left(0, \delta_2 d_2 - \frac{(1-\rho_2)}{(1-\rho)} \delta d_2 \right),$$

implying that the boundaries $\mathcal{B}1$ and $\mathcal{B}2$ have no windows. Therefore, the optimal surplus trajectory will keep bouncing back and forth between the two boundaries and never reach the cyclic schedule in finite time. In other words, the cyclic schedule can only be reached asymptotically when $c_1^* \to \infty$ and $c_2^* \to \infty$ (Figure 4.20 illustrates this case).
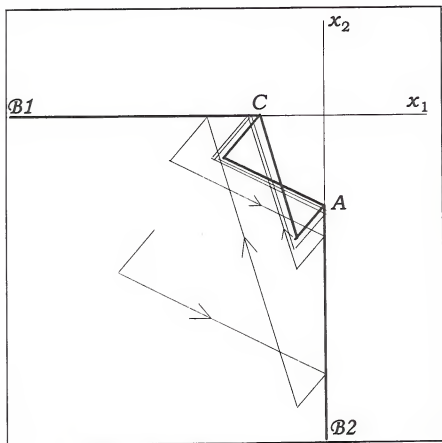

Figure 4.20: Case where $c_1^* \to \infty$, $c_2^* \to \infty$.

4.7 Summary

In this chapter, we studied the transient behavior of a deterministic one-machine two-product manufacturing system with setup times and no setup costs. We divided the surplus/backlog space into two major Regions $\Re^o$ and $\Re^u$. For initial surplus levels in Region $\Re^o$, the optimal solution was obtained by inspection. For initial surplus levels in Region $\Re^u$, two approaches were adopted: an algorithmic approach and an analytical one. For the algorithmic approach, a novel algorithm, to obtain the optimal state trajectory, was developed. The algorithm involves only a line search procedure, which makes it very fast. The algorithmic approach is useful for extending this work to the multi-part-type case. The analytical approach consists of the explicit expression of the optimal switching policy in Region $\Re^u$. The analytical approach is useful for studying special cases and determining the pertinent parameters of the system. Besides, it is the perfect real-time scheduling tool. The complete optimal transient solution is a feedback control policy.

CHAPTER 5
OPTIMAL TRANSIENT SOLUTION OF A TWO-PART-TYPE MANUFACTURING SYSTEM WITH SETUP TIMES AND SETUP COSTS DETERMINISTIC SYSTEM

5.1 Introduction

In this chapter, we extend the results of the previous chapter to the case where setup costs are nonnegligible anymore. In Section 5.2, we determine the optimal location in x-space of what we call the extracted cyclic schedule. In Section 5.3, we determine the optimal transient solution in Region $\Re^\circ$. In Section 5.4, we determine the optimal transient solution in Region $\Re^u$. In Section 5.5, we study the case where setup times are negligible. We conclude this chapter with a summary of results.

5.2 Optimal Location of the Extracted Cyclic Schedule

In Chapter 4, we studied the transient behavior of a system with zero setup costs. For a heavily loaded system (i.e., ρ is close to one), the condition $\rho + \rho_1 - 1 \geq 0$ holds and hence, τ_1^* and τ_2^* are equal to zero (see Theorem 3.2 of Chapter 3). In this case, there is no production at the demand rates for both part types over the cyclic schedule (Figure 3.3a of Chapter 3). In order to apply the results to the more general case of nonnegligible setup times and costs, we need to extract a schedule from the cyclic schedule of Figure 3.3d of Chapter 3 (which represents the general case). The extracted schedule has exactly the solution structure shown in Figure 3.3a. Figure 5.1 shows how to extract this schedule.
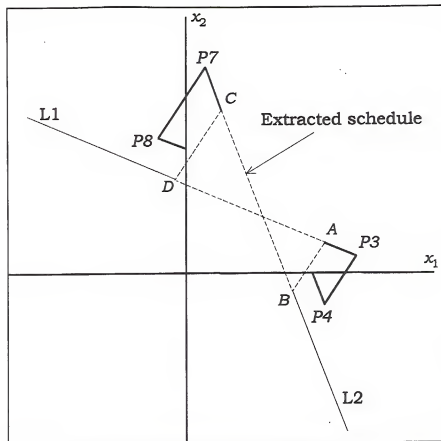
113

Figure 5.1: The extracted schedule.

Now, we determine the location of the extracted schedule. The latter is characterized by the coordinates of the points A, B, C, and D in x-space (Figure 5.1).

Let $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$, $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$, $C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$, and $D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}$.

As in Chapter 4, we denote by L1, the line containing Points D and $P3$ and by L2, the line containing Points B and $P7$ (Figure 5.1). L1 and L2 are given as follows:

Line L1: $d_2 \rho_1 (x_1 - P_{31}) + d_1 (1 - \rho_1)(x_2 - P_{32}) = 0;$ (5.1)

Line L2: $d_2 (1 - \rho_2)(x_1 - P_{71}) + d_1 \rho_2 (x_2 - P_{72}) = 0.$ (5.2)

Then, we have

$$B = A - \begin{pmatrix} \delta_2 d_1 \\ \delta_2 d_2 \end{pmatrix}, \; D = C - \begin{pmatrix} \delta_1 d_1 \\ \delta_1 d_2 \end{pmatrix}, \; A \in L1, \; B \in L2, \; C \in L2, \text{ and } D \in L1.$$

Hence, the coordinates of Point A and Point C can be determined by solving the following system of linear equations, respectively:

$$\begin{cases} d_2 \rho_1 (A_1 - P_{31}) + d_1 (1 - \rho_1)(A_2 - P_{32}) = 0, \\ d_2 (1 - \rho_2)(A_1 - \delta_2 d_1 - P_{71}) + d_1 \rho_2 (A_2 - \delta_2 d_2 - P_{72}) = 0; \end{cases}$$

$$\begin{cases} d_2 \rho_1 (C_1 - \delta_1 d_1 - P_{31}) + d_1 (1 - \rho_1)(C_2 - \delta_1 d_2 - P_{32}) = 0, \\ d_2 (1 - \rho_2)(C_1 - P_{71}) + d_1 \rho_2 (C_2 - P_{72}) = 0. \end{cases}$$

The coordinates of the extracted schedule are then given by

$$A = \begin{pmatrix} S_1 + (d_1/d_2)\rho_2 \alpha_1 Q_2 - \alpha_1 (1 - \rho_2)(Q_1 - \delta d_1) \\ s_2 + \delta_2 d_2 + (d_2/d_1)\rho_1 \alpha_2 (Q_1 - \delta d_1) - Q_2 \rho_1 \rho_2/(1 - \rho) \end{pmatrix}; \tag{5.3}$$

$$B = \begin{pmatrix} S_1 - \delta_2 d_1 + (d_1/d_2)\rho_2 \alpha_1 Q_2 - \alpha_1 (1 - \rho_2)(Q_1 - \delta d_1) \\ s_2 + (d_2/d_1)\rho_1 \alpha_2 (Q_1 - \delta d_1) - Q_2 \rho_1 \rho_2/(1 - \rho) \end{pmatrix}; \tag{5.4}$$

$$C = \begin{pmatrix} s_1 + \delta_1 d_1 + (d_1/d_2)\rho_2 \alpha_1 (Q_2 - \delta d_2) - Q_1 \rho_1 \rho_2/(1 - \rho) \\ S_2 + (d_2/d_1)\rho_1 \alpha_2 Q_1 - \alpha_2 (1 - \rho_1)(Q_2 - \delta d_2) \end{pmatrix}; \tag{5.5}$$

$$D = \begin{pmatrix} s_1 + (d_1/d_2)\rho_2 \alpha_1 (Q_2 - \delta d_2) - Q_1 \rho_1 \rho_2/(1 - \rho) \\ S_2 - \delta_1 d_2 + (d_2/d_1)\rho_1 \alpha_2 Q_1 - \alpha_2 (1 - \rho_1)(Q_2 - \delta d_2) \end{pmatrix}, \tag{5.6}$$

where s_i, S_i and Q_i are the optimal steady state parameters (Chapter 3), and

$$\alpha_i = \frac{(1 - \rho_i)}{(1 - \rho)} \; (i = 1, 2).$$

Notice that, if there is no production at the demand rate for both parts (i.e., $\tau_1 = 0$ and $\tau_2 = 0$), then $Q_1 = q_1$ and $Q_2 = q_2$. In this case, we recover the coordinates of the cyclic schedule studied in Chapter 4 which are given by

$$A = \begin{pmatrix} S_1 \\ s_2 + \delta_2 d_2 \end{pmatrix}, \; B = \begin{pmatrix} S_1 - \delta_2 d_1 \\ s_2 \end{pmatrix}, \; C = \begin{pmatrix} s_1 + \delta_1 d_1 \\ S_2 \end{pmatrix} \text{ and } D = \begin{pmatrix} s_1 \\ S_2 - \delta_1 d_2 \end{pmatrix}.$$

5.3 Partition of the x-Space into Two Mutually Exclusive Regions

To obtain the optimal solution of the transient problem, we partition the x-space into two mutually exclusive major regions $\mathfrak{R}^u$ and $\mathfrak{R}^o$ as shown in Figure 5.2. $\mathfrak{R}^u$ and $\mathfrak{R}^o$ are defined as in Chapter 4, but this time we use the coordinates of the extracted schedule.

As in Chapter 4, we carry out the transient analysis in two parts. In the first part, we determine the optimal transient solution for all initial surplus levels in region $\mathfrak{R}^o$. In the second part, the optimal transient solution for initial surplus levels in Region $\mathfrak{R}^u$ is derived.

Without loss of generality, we index the parts such that Part Type 1 is the part type with the larger setup time and setup cost (i.e., $\delta_1 \geq \delta_2$ and $k_1 \geq k_2$).
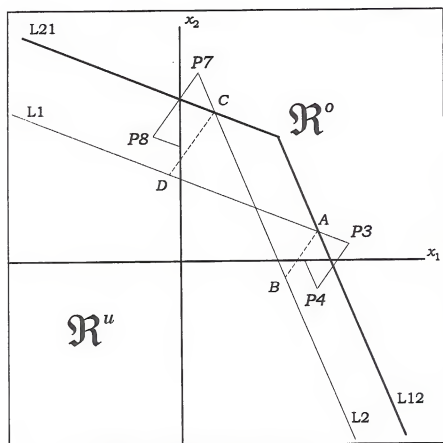

Figure 5.2: Regions $\mathfrak{R}^u$ and $\mathfrak{R}^o$.

5.4 Optimal Transient Solution in Region $\mathfrak{R}^\circ$

To determine the optimal solution for initial surplus levels in Region $\mathfrak{R}^\circ$, we apply the same technique used in Chapter 4. First, we partition Region $\mathfrak{R}^\circ$ into three mutually exclusive regions, G, G1 and G2. Then, we further partition Region G into eight mutually exclusive regions G11, G12, G21, G22, H11, H12, H21, and H22. Figures 5.3 and 5.4 show the partition of Region $\mathfrak{R}^\circ$ and G respectively. These regions are defined as follows:

$$G11 = \{(x_1, x_2) | x_1 - \delta_1 d_1 \geq 0; \ -d_2(x_1 - P_{71}) + d_1(x_2 - P_{72}) \geq 0\};$$

$$G12 = \{(x_1, x_2) | d_2(x_1 - P_{71}) - d_1(x_2 - P_{72}) > 0;$$
$$-d_2(x_1 - E_1) + d_1(x_2 - E_2) \geq 0;$$
$$d_2(1 - \rho_2)(x_1 - A_1) + d_1 \rho_2(x_2 - A_2) \geq 0\};$$

$$G21 = \{(x_1, x_2) | -d_2(x_1 - P_{31}) + d_1(x_2 - P_{32}) \geq 0;$$
$$d_2(x_1 - E_1) - d_1(x_2 - E_2) > 0;$$
$$d_2 \rho_1(x_1 - C_1) + d_1(1 - \rho_1)(x_2 - C_2) \geq 0\};$$

$$G22 = \{(x_1, x_2) | x_2 - \delta_2 d_2 \geq 0; \ d_2(x_1 - P_{31}) - d_1(x_2 - P_{32}) > 0\};$$

$$H11 = \{(x_1, x_2) | x_1 - \delta_1 d_1 \geq 0;$$
$$d_2(x_1 - P_{71}) - d_1(x_2 - P_{72}) > 0;$$
$$-d_2(x_1 - g_{11}) + d_1(x_2 - g_{12}) \geq 0;$$
$$-d_2(1 - \rho_2)(x_1 - A_1) - d_1 \rho_2(x_2 - A_2) \geq 0\};$$

$$H12 = \{(x_1, x_2) | d_2 \rho_1(x_1 - C_1) - d_1(1 - \rho_1)(x_2 - C_2) > 0;$$
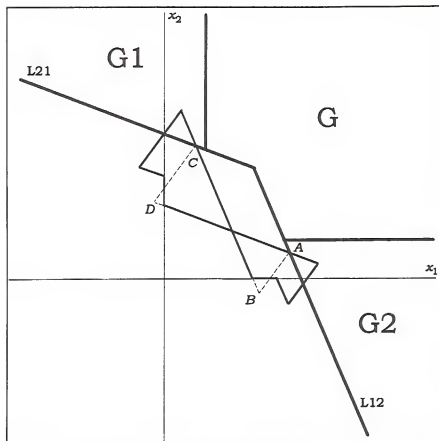$$-d_2(x_1 - g_{11}) + d_1(x_2 - g_{12}) \geq 0;$$
$$-d_2(1 - \rho_2)(x_1 - A_1) - d_1 \rho_2(x_2 - A_2) \geq 0\};$$

$$H21 = \{(x_1, x_2) | d_2(x_1 - I_1) - d_1(x_2 - I_2) > 0;$$
$$d_2(1 - \rho_2)(x_1 - A_1) - d_1 \rho_2(x_2 - A_2) \geq 0;$$
$$-d_2 \rho_1(x_1 - C_1) - d_1(1 - \rho_1)(x_2 - C_2) > 0;$$
$$-d_2(x_1 - g_{21}) + d_1(x_2 - g_{22}) \geq 0\};$$

Figure 5.3: Partition of Region $\Re^\circ$.

$$\text{H22} = \{(x_1, x_2) | x_2 - \delta_2 d_2 \geq 0;$$

$$-d_2(x_1 - P_{31}) - d_1(x_2 - P_{32}) > 0;$$

$$-d_2\rho_1(x_1 - C_1) - d_1(1 - \rho_1)(x_2 - C_2) > 0;$$

$$-d_2(x_1 - g_{21}) + d_1(x_2 - g_{22}) > 0\};$$

$$\text{G1} = \{(x_1, x_2) | -x_1 + \delta_1 d_1 > 0; \ d_2\rho_1(x_1 - P_{71}) + d_1(1 - \rho_1)(x_2 - P_{72}) \geq 0\};$$

$$\text{G2} = \{(x_1, x_2) | -x_2 + \delta_2 d_2 > 0; \ d_2(1 - \rho_2)(x_1 - P_{31}) + d_1\rho_2(x_2 - P_{32}) \geq 0\},$$

where $g_1 = (g_{11}, g_{12})^T$ is the point in x-space given by

$$g_{11} = \delta_1 d_1 \text{ and } g_{12} = C_2 + (C_1 - \delta_1 d_1) d_2 \rho_1 / d_1(1 - \rho_1),$$

and $g_2 = (g_{21}, g_{22})^T$ is the point in x-space given by

$$g_{21} = A_1 + (A_2 - \delta_2 d_2) d_1 \rho_2 / d_2(1 - \rho_2) \text{ and } g_{22} = \delta_2 d_2.$$

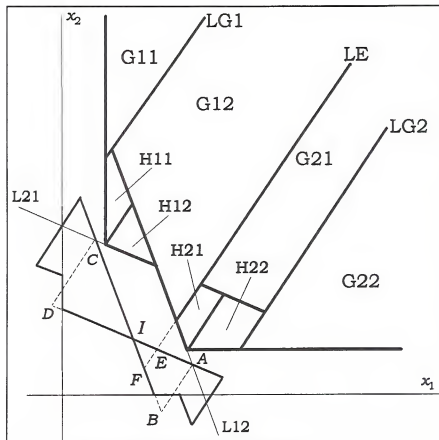$E = (E_1, E_2)^T$ is the point in x-space given by

Figure 5.4: Partition of Region G.

$$E_1 = \left(-\theta_2 + \sqrt{\theta_2^2 + \theta_1\theta_3}\right)\Big/\theta_1;$$

$$E_2 = \xi_1\left(-\theta_2 + \sqrt{\theta_2^2 + \theta_1\theta_3}\right)\Big/\theta_1 + \xi_2,$$

where

$$\theta_1 = \left(c_1^+/d_1\right)\left(\xi_1^2 - \xi_3^2\right) + \left(c_2^+/d_2\right)\left(1 - \zeta_5^2\right);$$

$$\theta_2 = \left(c_1^+/d_1\right)\left(\xi_1\xi_2 - \xi_3\xi_4\right) - \left(c_2^+/d_2\right)\zeta_5\zeta_6;$$

$$\theta_3 = 2\left(k_1 - k_2\right) + \left(c_1^+/d_1\right)\left(\xi_4^2 - \xi_2^2\right) + \left(c_2^+/d_2\right)\zeta_6^2;$$

$$\xi_1 = -\left(d_1/d_2\right)\left(1 - \rho_1\right)\Big/\rho_1;$$

$$\xi_2 = A_1 + A_2\left(d_1/d_2\right)\left(1 - \rho_1\right)\Big/\rho_1;$$

$$\xi_3 = -\left(d_1/d_2\right)\left(\rho_2/\rho_1\right);$$

$$\xi_4 = \rho_2 A_1 + \left(\rho_2/\rho_1\right)\left(1 - \rho_1\right)A_2 + \left(1 - \rho_2\right)C_1 + \rho_2\left(d_1/d_2\right)C_2;$$

$$\xi_5 = \left(1 - \rho_2\right)\Big/\rho_1;$$

$$\xi_6 = \rho_2 C_2 - A_2 (1 - \rho_2)(1 - \rho_1)/\rho_1 + (d_2/d_1)(1 - \rho_2)(C_1 - A_1).$$

Remark: Line LE is the boundary in x-space separating Region G12 and G21. Since the direction of Line LE is known (given by the vector $(-d_1, -d_2)$ in x-space), it is sufficient to find the coordinates of Point E in x-space to completely define Boundary LE. To determine Point E, we note that for initial surplus levels belonging to G12∩G21 (i.e., on Line LE), the cost of reaching the cyclic schedule at Point E with a setup for Part Type 2 and the cost of reaching the cyclic schedule at Point F with a setup for Part Type 1 must be equal. Mathematically, this can be written as follows:

$$k_1 + J = k_2 + J + (c_1^*/d_1)(E_1^2 - F_1^2) + (c_2^*/d_2)(E_2^2 - F_2^2),$$

where J represents the cost of reaching Point E with a setup for either part type.

Noticing that $E \in L1 \cap LE$ and $F \in L2 \cap LE$, we express F_1, F_2, and E_1 as a function of E_2 and then substitute in the above cost equality. The result is a quadratic in E_2. E_1 is then obtained by substituting E_2 by its expression from the quadratic solution as shown above.

If $E_1 > A_1$, Region G21 becomes empty (G21 = ∅). This case corresponds to a very high setup cost of Part Type 1 compared to that of Part Type 2. Therefore it is more economical to setup the machine for Part Type 2 before the cyclic schedule is reached.

In the case of equal setup costs for both part types, Point E coincides with Point I. In this case, we obtain the exact partition obtained in Chapter 4 for the case of no setup costs.

To summarize the control actions in Region $\Re^c$, let $x = (a, b)$ be the vector of initial surplus levels.

- If $x \in$ G1:

  1: Setup the machine for Part Type 1;

  2: Produce Part Type 1 at the rate U_1;

  3: When the surplus level of Part Type 1 becomes 0, change the production rate to d_1;

  4: When the surplus level of Part Type 2 becomes $x_2 = A_2 + A_1 d_2 \rho_1 / d_1 (1 - \rho_1)$, switch to the control actions of the cyclic schedule.

- If $x \in$ G11∪H11:

  1: Do not produce either part type;

  2: When the surplus level of Part Type 1 becomes $\delta_1 d_1$, start a setup for Part Type 1;

  3: When the setup is over, produce Part Type 1 at the demand rate d_1;

  4: When the surplus level of Part Type 2 becomes $x_2 = A_2 + A_1 d_2 \rho_1 / d_1 (1 - \rho_1)$, switch to the control actions of the cyclic schedule.

- If $x \in$ G12∪H21:

  1: Do not produce either part type;

  2: When the surplus level of Part Type 2 reaches Level l_2, immediately start a setup for Part Type 2. l_2 is given by $l_2 = (1 - \rho_2)(b - (d_2/d_1)a) + (d_2/d_1)(1 - \rho_2)A_1 + \rho_2 A_2$;

  3: When the setup is over, switch to the cyclic schedule control actions.

- If $x \in$ G21∪H12:

  1: Do not produce either part type;

2: When the surplus level of Part Type 1 reaches Level l_1, immediately start a setup for Part Type 1. l_1 is given by

$$l_1 = \rho_1(b - (d_2/d_1)a) + (d_2/d_1)\rho_1 C_1 + (1 - \rho_1)C_2;$$

3: When the setup is over, switch to the cyclic schedule control actions.

- If $x \in \text{G22} \cup \text{H22}$:

  1: Do not produce either part type;

  2: When the surplus level of Part Type 2 becomes $\delta_2 d_2$, start a setup for Part Type 2;

  3: When the setup is over, produce Part Type 2 at the demand rate d_2;

  4: When the surplus level of Part Type 1 becomes $x_1 = C_1 + C_2 d_1 \rho_2 / d_2(1 - \rho_2)$, switch to the control actions of the cyclic schedule.

- If $x \in \text{G2}$:

  1: Setup the machine for Part Type 2;

  2: When the setup is over, produce Part Type 2 at the rate U_2;

  3: When the surplus level of Part Type 2 becomes 0, change the production rate to d_2;

  4: When the surplus level of Part Type 1 becomes $x_1 = C_1 + C_2 d_1 \rho_2 / d_2(1 - \rho_2)$, switch to the control actions of the cyclic schedule.

5.5 Optimal Transient Solution in Region $\Re^u$

As has been shown in Chapter 4, the optimal transient solution in Region $\Re^u$ is characterized by the optimal setup switching policy in that region,

since we know that the optimal production rates in Region $\mathfrak{R}^u$ are the maximum production rates for both part types. In addition, in Chapter 4, it has been established that the optimal switching policy is a special corridor policy with two windows. In the case where production at the demand rates is possible in the cyclic schedule, this result is still valid and in fact is optimal when applied to the extracted schedule characterized by its location in x-space given by A, B, C, and D. Figure 5.5 illustrates this corridor policy which is completely characterized by its corridor walls defined by boundaries $\mathcal{B}1$ and $\mathcal{B}2$ and its corridor windows defined by the pairs of points $(W1, V1)$ and $(W2, V2)$. Using the technique developed in Chapter 4 (Section 4.6), we can determine the corridor boundaries and windows as follows:
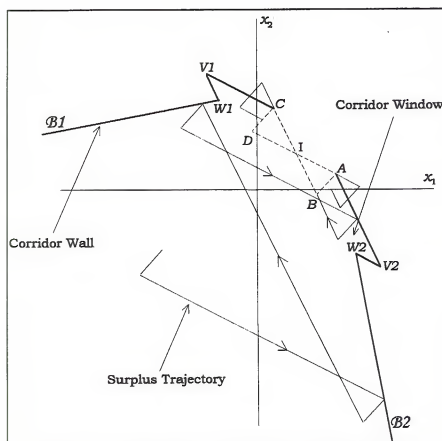

Figure 5.5: Illustration of the setup switching policy.

5.5.1 Equations of Boundaries $\mathcal{B}1$ and $\mathcal{B}2$

$$\mathcal{B}1: \; M_1 x_1 + M_2 x_2 + M_3 = 0, \tag{5.7}$$

where

$$M_1 = c_1^- / (1 - \rho_1) - \rho_1 \rho_2 c_1^* / (1 - \rho_1)(1 - \rho) - (d_1/d_2)(\rho_1/\rho_2)\alpha_2 c_2^-; \tag{5.8a}$$

$$M_2 = -(d_1/d_2)\rho_2 c_2^* / (1 - \rho) - c_2^* / \rho_2 - \alpha_1 (1 - \rho_2) c_2^- / \rho_2; \tag{5.8b}$$

$$M_3 = -(M_1 C_1 + M_2 C_2). \tag{5.8c}$$

$$\mathcal{B}2: \; N_1 x_1 + N_2 x_2 + N_3 = 0, \tag{5.9}$$

where

$$N_1 = -(d_2/d_1)\rho_1 c_1^* / (1 - \rho) - c_1^* / \rho_1 - \alpha_1 (1 - \rho_2) c_1^- / \rho_1; \tag{5.10a}$$

$$N_2 = c_2^- / (1 - \rho_2) - \rho_1 \rho_2 c_2^* / (1 - \rho_2)(1 - \rho) - (d_2/d_1)(\rho_2/\rho_1)\alpha_1 c_1^-; \tag{5.10b}$$

$$N_3 = -(N_1 A_1 + N_2 A_2). \tag{5.10c}$$

5.5.2 Corridor windows

As in Chapter 4, Let $W1 = (w_{11}, w_{12})$, $V1 = (v_{11}, v_{12})$, $W2 = (w_{21}, w_{22})$, and $V2 = (v_{21}, v_{22})$ in x-space. Then, $W1$ satisfies the following system:

$$\begin{cases} a_1 w_{11}^2 + a_2 w_{11} + a_3 = 0; \\ w_{12} = -(M_1/M_2)w_{11} - (M_3/M_2); \end{cases} \tag{5.11}$$

$$\begin{cases} v_{11} = w_{11} + \rho_1 \rho_2 (w_{11} - C_1)/(1-\rho) + (d_1/d_2)\rho_2 (1-\rho_1)(w_{12} - C_2)/(1-\rho); \\ v_{12} = C_2 + \rho_1 \rho_2 (C_2 - w_{12})/(1-\rho) + (d_2/d_1)\rho_1 (1-\rho_2)(C_1 - w_{11})/(1-\rho) \end{cases} \tag{5.12}$$

where

$$a_1 = \frac{(\gamma_{11}^2 - 1)c_1^- - \eta_{11}^2 c_1^*}{2d_1(1-\rho_1)} + \frac{(\gamma_{12}^2 - (M_1/M_2)^2)c_2^* - \eta_{12}^2 c_2^-}{2d_2(1-\rho_2)}; \tag{5.13a}$$

$$a_2 = \frac{(\gamma_{11}(\mu_{11} - \delta_1 d_1) + \delta_1 d_1)c_1^- - \eta_{11}\nu_{11}c_1^*}{d_1(1-\rho_1)}$$

$$+\frac{\left(\gamma_{12}\mu_{12}-M_1M_3/M_2^2\right)c_2^+-\eta_{12}(\nu_{12}-\delta_2 d_2)c_2^-}{d_2(1-\rho_2)};\tag{5.13b}$$

$$a_3=\frac{\mu_{11}(\mu_{11}-2\delta_1 d_1)c_1^--(\nu_{11}^2-I_1^2)c_1^+}{2d_1(1-\rho_1)}$$

$$+\frac{\left(\mu_{12}^2-(M_3/M_2)^2-I_2^2\right)c_2^+-(\nu_{12}-\delta_2 d_2)^2c_2^-}{2d_2(1-\rho_2)}-k_2\tag{5.13c}$$

$$\eta_{11}=\rho_2(d_1/d_2)(M_1/M_2)\alpha_1-\rho_1/(1-\rho);\tag{5.14a}$$

$$\eta_{12}=\alpha_2\{(d_2/d_1)\rho_1-(M_1/M_2)(1-\rho_1)\};\tag{5.14b}$$

$$\nu_{11}=\alpha_1(1-\rho_2)A_1+\alpha_1\rho_2(d_1/d_2)\left(A_2+(M_3/M_2)\right)+\delta_1 d_1\rho_2/(1-\rho);\tag{5.15a}$$

$$\nu_{12}=-\rho_1\alpha_2(d_2/d_1)A_1-\rho_1\rho_2 A_2/(1-\rho)-(1-\rho_2)\alpha_1(M_3/M_2)-\alpha_2(d_2/d_1)\delta_1 d_1;\tag{5.15b}$$

$$\gamma_{11}=\alpha_1\{(1-\rho_2)-\rho_2(d_1/d_2)(M_1/M_2)\};\tag{5.16a}$$

$$\gamma_{12}=\rho_1\rho_2(M_1/M_2)/(1-\rho)-\rho_1\alpha_2(d_2/d_1);\tag{5.16b}$$

$$\mu_{11}=-\rho_1\rho_2 C_1/(1-\rho)-\rho_2\alpha_1(d_1/d_2)\left(C_2+(M_3/M_2)\right);\tag{5.17a}$$

$$\mu_{12}=\rho_1\alpha_2(d_2/d_1)C_1+(1-\rho_1)\alpha_2 C_2+\rho_1\rho_2(M_3/M_2)/(1-\rho);\tag{5.17b}$$

$W2$ satisfies the following system:

$$\begin{cases}b_1 w_{22}^2+b_2 w_{22}+b_3=0;\\ w_{21}=-(N_2/N_1)w_{22}-(N_3/N_1);\end{cases}\tag{5.18}$$

$$\begin{cases}v_{21}=A_1+\rho_1\rho_2(A_1-w_{21})/(1-\rho)+(d_1/d_2)\rho_2(1-\rho_1)(A_2-w_{22})/(1-\rho);\\ v_{22}=w_{22}+\rho_1\rho_2(w_{22}-A_2)/(1-\rho)+(d_2/d_1)\rho_1(1-\rho_2)(w_{21}-A_1)/(1-\rho)\end{cases}\tag{5.19}$$

where

$$b_1=\frac{\left(\gamma_{21}^2-(N_2/N_1)^2\right)c_1^+-\eta_{21}^2 c_1^-}{2d_1(1-\rho_1)}+\frac{(\gamma_{22}^2-1)c_2^--\eta_{22}^2 c_2^+}{2d_2(1-\rho_2)};\tag{5.20a}$$

$$b_2=\frac{\left(\gamma_{21}\mu_{21}-N_2 N_3/N_1^2\right)c_1^+-\eta_{21}(\nu_{21}-\delta_1 d_1)c_1^-}{d_1(1-\rho_1)}$$

$$+\frac{\left(\gamma_{22}(\mu_{22}-\delta_2 d_2)+\delta_2 d_2\right)c_2^--\eta_{22}\nu_{22}c_2^+}{d_2(1-\rho_2)};\tag{5.20b}$$

$$b_3 = \frac{(\mu_{21}^2 - (N_3/N_1)^2 - I_1^2)c_1^+ - (\nu_{21} - \delta_1 d_1)^2 c_1^-}{2d_1(1 - \rho_1)}$$

$$+ \frac{\mu_{22}(\mu_{22} - 2\delta_2 d_2)c_2^- - (\nu_{22}^2 - I_2^2)c_2^+}{2d_2(1 - \rho_2)} - k_1; \qquad (5.20c)$$

$$\eta_{21} = \alpha_1\{(d_1/d_2)\rho_2 - (N_2/N_1)(1 - \rho_2)\}; \qquad (5.21a)$$

$$\eta_{22} = \rho_1\alpha_2(d_2/d_1)(N_2/N_1) - \rho_1\rho_2/(1 - \rho); \qquad (5.21b)$$

$$\nu_{21} = -\rho_1\rho_2 C_1/(1 - \rho) - \rho_2\alpha_1(d_1/d_2)C_2 - (1 - \rho_2)\alpha_1(N_3/N_1) - \alpha_1(d_1/d_2)\delta_2 d_2; \qquad (5.22a)$$

$$\nu_{22} = \rho_1\alpha_2(d_2/d_1)(C_1 + (N_3/N_1)) + (1 - \rho_1)\alpha_2 C_2 + \delta_2 d_2\rho_1/(1 - \rho); \qquad (5.22b)$$

$$\gamma_{21} = \rho_1\rho_2(N_2/N_1)/(1 - \rho) - \rho_2\alpha_1(d_1/d_2); \qquad (5.23a)$$

$$\gamma_{22} = \alpha_2\{(1 - \rho_1) - \rho_1(d_2/d_1)(N_2/N_1)\}; \qquad (5.23b)$$

$$\mu_{21} = (1 - \rho_2)\alpha_1 A_1 + \rho_2\alpha_1(d_1/d_2)A_2 + \rho_1\rho_2(N_3/N_1)/(1 - \rho); \qquad (5.24a)$$

$$\mu_{22} = -\rho_1\alpha_2(d_2/d_1)(A_1 + (N_3/N_1)) - \rho_1\rho_2 A_2/(1 - \rho). \qquad (5.24b)$$

Notice that in (5.11), w_{11} must be a real root of the quadratic such that $w_{11} \le C_1$. Similarly, w_{22} in (5.18) must be a real root such that $w_{22} \le A_2$. If either quadratic in (5.11) or (5.18) has no real root, the corridor window does not exist. Which means that, it is always optimal to reach the extracted schedule from the other side (i.e., the side with the corridor window).

The optimal trajectory emanating from a point $X=(a,b)$ in Region $\Re^u$ is obtained by using Procedure 1 (Section 4.6).

Compared to the case of zero setup costs, the only change is in the expressions of a_3 and b_3 given above by (5.13c) and (5.20c) respectively, where $-k_2$ was added to a_3 and $-k_1$ was added to b_3 originally obtained in Chapter 4. This result is intuitive, since in the case of nonzero setup costs, we have a wider corridor windows, which means that if the setup costs are high then it might not be optimal to add another setup change and therefore reach the extracted schedule producing the same part type that the machine is set up

for, avoiding a rather expensive setup. Also, note that the setup costs do not appear explicitly in the equations of Boundaries $\mathcal{B}1$ and $\mathcal{B}2$, but the latter depend on the setup costs implicitly through the coordinates of Points A and C of the extracted schedule.

5.5.3 Numerical example

Consider the following system:

Part Type 1:
$c_1^+ = \$2/\text{unit/day}$, $c_1^- = \$10/\text{unit/day}$, $U_1 = 8$ unit/day,
$d_1 = 2.5$ unit/day, $\delta_1 = 1$ day, and $k_1 = \$100$;

Part Type 2:
$c_2^+ = \$2/\text{unit/day}$, $c_2^- = \$10/\text{unit/day}$, $U_2 = 10$ unit/day,
$d_2 = 3$ unit/day, $\delta_2 = 1$ day, and $k_2 = \$120$.

Figure 5.6 shows the optimal location and shape of the cyclic schedule as well as the optimal trajectory emanating at Point (-20,-20) (i.e., we start 20 units short of both part types) in Region $\mathfrak{R}^u$, as well as Boundaries $\mathcal{B}1$ and $\mathcal{B}2$ and the two pairs $(W1, V1)$ and $(W2, V2)$ defining the two corridor windows. The optimal production and setup planning for this system can be described as follows (after rounding to integers): Set up the machine and produce Part Type 2 at the rate of 10 units/day until its surplus level reaches 14 units (i.e., when the surplus trajectory hits Boundary $\mathcal{B}1$); set up the machine and produce Part Type 1 at the rate of 8 units/day until its surplus level reaches 10 units (i.e., when the surplus trajectory hits Boundary $\mathcal{B}2$), set up the machine and produce Part Type 2 at the rate of 10 units/day until its surplus level reaches 15 units (i.e., when the surplus trajectory enters Window $(W1, V1)$); set up the machine and produce Part Type 1 at the rate of 8 units/day until its surplus

level reaches 16 units (which corresponds to Point *P3* on the cyclic schedule); switch to the control actions of the cyclic schedule given as follows: Set up the machine and produce Part Type 2 at the rate of 10 units/day until its surplus level reaches 0, continue producing Part Type 2 at the demand rate of 3 units/day until the surplus level of Part Type 1 drops to 5 units; increase the production rate of Part Type 2 to 10 units/day until its surplus level reaches 16 units; set up the machine and produce Part Type 1 at the rate of 8 units/day until its surplus level reaches 0; continue producing Part Type 1 at the demand rate of 2.5 units/day until the surplus level of Part Type 2 drops to 8 units; increase the production rate of Part Type 1 to 8 units/day until its surplus level reaches 16 units, which is the point we started the cyclic schedule at.

5.6 Case of Zero Setup Times and Nonzero Setup Costs

In this section, we consider the case where setup times are very small or are order of magnitude smaller than the other parameters of the system (especially setup costs). This kind of situation may arise when the setups can be accomplished relatively quickly, but the equipment, the material (cleaning products for instance), and the labor used for setups are very expensive.

5.6.1 Optimal location of the extracted cyclic schedule

Theorem 3.3 of Chapter 3 states that it is not possible to have simultaneously no production at the demand rate for both part types, in the case of zero setup times. Therefore, the general structure of the cyclic schedule is as shown in Figure 5.7. The coordinates of the cyclic schedule in this case are given as follows:
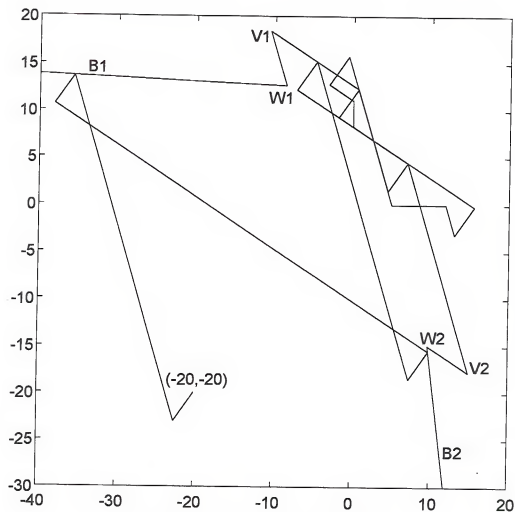
Figure 5.6: Example illustration

$$P1 = \begin{pmatrix} 0 \\ S_2 + s_1(d_2/d_1)\rho_1/(1-\rho_1) \end{pmatrix};$$

$$P2 = \begin{pmatrix} 0 \\ S_2 + s_1(d_2/d_1)\rho_1/(1-\rho_1) - \tau_1 d_2 \end{pmatrix};$$

$$P3 = P4 = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix};$$

$$P5 = \begin{pmatrix} S_1 + s_2(d_1/d_2)\rho_2/(1-\rho_2) \\ 0 \end{pmatrix};$$
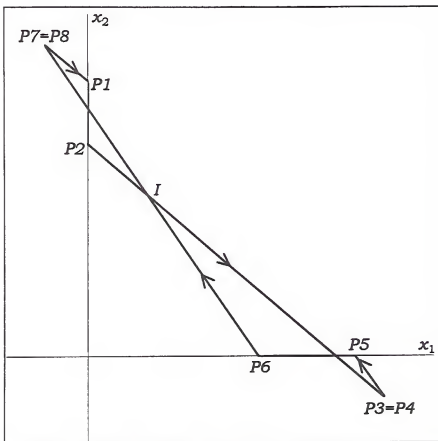
Figure 5.7: General structure of the cyclic schedule
in the case of no setup times.

$$P6 = \begin{pmatrix} S_1 + s_2(d_1/d_2)\rho_2/(1-\rho_2) - \tau_2 d_1 \\ 0 \end{pmatrix};$$

$$P7 = P8 = \begin{pmatrix} s_1 \\ S_2 \end{pmatrix},$$

where s_i, S_i, Q_i and τ_i $(i=1,2)$ are the optimal steady state parameters.

The extracted schedule in this case reduces to a single point in x-space (Point I in Figure 5.7). To see this, let $\delta_1 = \delta_2 = 0$ in the coordinates of the extracted schedule calculated for the nonzero setup times case (Equations (5.3)-(5.6)). This gives

$$A = B = \begin{pmatrix} S_1 + (d_1/d_2)\rho_2\alpha_1 Q_2 - (1-\rho_2)\alpha_1 Q_1 \\ s_2 + (d_2/d_1)\rho_1\alpha_2 Q_1 - \rho_1\rho_2 Q_2/(1-\rho) \end{pmatrix};$$

$$C = D = \begin{pmatrix} s_1 + (d_1/d_2)\rho_2\alpha_1 Q_2 - \rho_1\rho_2 Q_1/(1-\rho) \\ S_2 + (d_2/d_1)\rho_1\alpha_2 Q_1 - (1-\rho_1)\alpha_2 Q_2 \end{pmatrix}.$$

Substituting $s_i = S_i - Q_i$ in the coordinates of A and B, gives Points C and D. Therefore,

$$A = B = C = D = I = \begin{pmatrix} S_1 + (d_1/d_2)\rho_2\alpha_1 Q_2 - (1-\rho_2)\alpha_1 Q_1 \\ s_2 + (d_2/d_1)\rho_1\alpha_2 Q_1 - \rho_1\rho_2 Q_2/(1-\rho) \end{pmatrix}.$$

5.6.2 Optimal transient solution

As for the case of nonzero setup times, we divide the x-space into two mutually exclusive major regions. Region $\mathfrak{R}^u$ and Region $\mathfrak{R}^o$ defined as follows:

$$\mathfrak{R}^u = \{(x_1, x_2) \mid d_2(1-\rho_2)(x_1 - I_1) + d_1\rho_2(x_2 - I_2) < 0;$$

$$d_2\rho_1(x_1 - I_1) + d_1(1-\rho_1)(x_2 - I_2) < 0\};$$

$$\mathfrak{R}^o = \{(x_1, x_2) \mid d_2(1-\rho_2)(x_1 - I_1) + d_1\rho_2(x_2 - I_2) \geq 0;$$

$$d_2\rho_1(x_1 - I_1) + d_1(1-\rho_1)(x_2 - I_2) \geq 0\}.$$

This partition of the x-space is shown in Figure 5.8. In this case, Lines L12 and L21 coincide with Lines L2 and L1, respectively. Lines L1 and L2 are defined as follows:

$$\text{Line L1: } d_2\rho_1(x_1 - I_1) + d_1(1-\rho_1)(x_2 - I_2) = 0;$$

$$\text{Line L2: } d_2(1-\rho_2)(x_1 - I_1) + d_1\rho_2(x_2 - I_2) = 0.$$

In the following, we study the optimal transient solution for initial surplus levels in Region $\mathfrak{R}^u$. The optimal transient solution for initial surplus levels in Region $\mathfrak{R}^o$ can be obtained by inspection in a similar way as in the case of nonzero setup times.
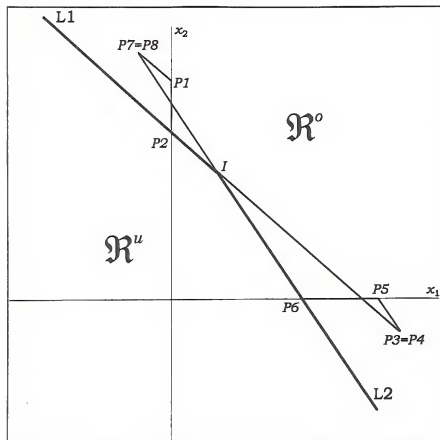
Figure 5.8: Partition of x-space in the case of zero setup times

Recall that the optimal switching policy in Region $\mathfrak{R}^u$ for the case of nonzero setup times is a special corridor policy characterized by boundaries $\mathcal{B}1$ and $\mathcal{B}2$, and the two windows defined by the pairs $(W1, V1)$ and $(W2, V2)$. In the case of zero setup times, to obtain the optimal policy, all we need to do is to let $\delta_1 = 0$ and $\delta_2 = 0$ in the equations of Boundaries $\mathcal{B}1$ and $\mathcal{B}2$ as well as in the coordinates of Points $W1$, $W2$, $V1$ and $V2$.

Equations of Boundaries $\mathcal{B}1$ and $\mathcal{B}2$

$$\mathcal{B}1: M_1 x_1 + M_2 x_2 + M_3 = 0;$$
$$\mathcal{B}2: N_1 x_1 + N_2 x_2 + N_3 = 0,$$

where M_1, M_2, N_1, and N_2 are as defined by (5.8a), (5.8b), (5.10a) and (5.10b) respectively, and $M_3 = -(M_1 I_1 + M_2 I_2)$, and $N_3 = -(N_1 I_1 + N_2 I_2)$.

<u>Corridor windows</u>

W1 and *V1* are given by

$$\begin{cases} a_1 w_{11}^2 + a_2 w_{11} + a_3 = 0; \\ w_{12} = -(M_1/M_2)w_{11} - (M_3/M_2); \end{cases}$$

$$\begin{cases} v_{11} = w_{11} + \rho_1\rho_2(w_{11} - C_1)/(1-\rho) + (d_1/d_2)\rho_2(1-\rho_1)(w_{12} - C_2)/(1-\rho); \\ v_{12} = C_2 + \rho_1\rho_2(C_2 - w_{12})/(1-\rho) + (d_2/d_1)\rho_1(1-\rho_2)(C_1 - w_{11})/(1-\rho); \end{cases}$$

W2 and *V2* are given by

$$\begin{cases} b_1 w_{22}^2 + b_2 w_{22} + b_3 = 0; \\ w_{21} = -(N_2/N_1)w_{22} - (N_3/N_1); \end{cases}$$

$$\begin{cases} v_{21} = A_1 + \rho_1\rho_2(A_1 - w_{21})/(1-\rho) + (d_1/d_2)\rho_2(1-\rho_1)(A_2 - w_{22})/(1-\rho); \\ v_{22} = w_{22} + \rho_1\rho_2(w_{22} - A_2)/(1-\rho) + (d_2/d_1)\rho_1(1-\rho_2)(w_{21} - A_1)/(1-\rho); \end{cases}$$

where, a_1, a_2, a_3, b_1, b_2, and b_3 are obtained by plugging $\delta_1 = 0$ and $\delta_2 = 0$ in (5.13a), (5.13b), (5.13c), (5.20a), (5.20b) and (5.20c) respectively.

<u>Numerical Example</u>

We consider the same example studied in Section 5.5 and let the setup times be zero and double the setup cost of each part type. The reason for increasing the setup costs for both part types can be thought of as using more sophisticated and costlier techniques, tools or labor to reduce the setup times to a nonsignificant level compared to the other parameters of the system.

Part Type 1:

$c_1^+ = \$2/\text{unit/day}$, $c_1^- = \$10/\text{unit/day}$, $U_1 = 8$ unit/day,

$d_1 = 2.5$ unit/day, $\delta_1 = 0$, and $k_1 = \$200$;

Part Type 2:

$c_2^+ = \$2/\text{unit/day}$, $c_2^- = \$10/\text{unit/day}$, $U_2 = 10$ unit/day,

$d_2 = 3$ unit/day, $\delta_2 = 0$, and $k_2 = \$240$.

The optimal trajectory and the optimal shape and location of the cyclic schedule are shown in Figure 5.9.
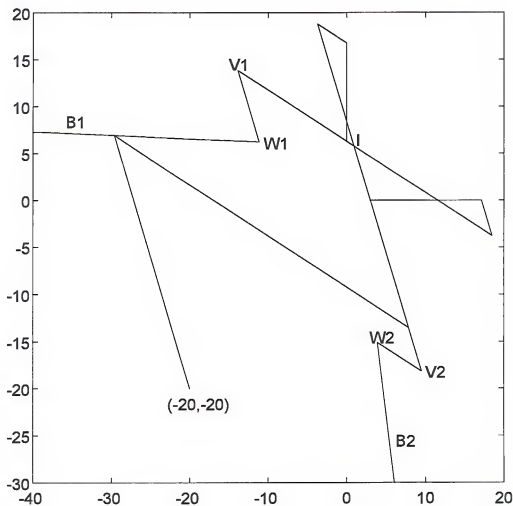

Figure 5.9: Example illustration

5.7 Summary

In this chapter, we extended the results of Chapter 4 to the more general case of nonzero setup times and nonzero setup costs. We showed that the optimal trajectory reaches the extracted schedule first, then follows the a cyclic schedule thereafter. Also, we studied the case where setup times can be

neglected compared to the remaining parameters of the system. In this case, we showed that the extracted schedule reduces to a single point in x-space, and obtained the optimal transient solution analytically as a special case of the more general case of nonzero setup times and costs.

CHAPTER 6
OPTIMAL AND NEAR OPTIMAL CONTROL
OF A TWO-PART-TYPE STOCHASTIC MANUFACTURING SYSTEM
WITH NONRESUMABLE SETUPS

6.1 Introduction

In this chapter, we consider the same system studied in the previous chapters with the following additional assumptions: The manufacturing system is unreliable, the setup times are now random and nonresumable, the planning horizon is infinite, and the optimization criterion is the total discounted cost over the planning horizon. The purpose is (*i*) to propose a stochastic control model for the simultaneous planning of production and setups in the unreliable manufacturing system; (*ii*) to develop an efficient technique for the computation of the optimal control policy; and (*iii*) to provide heuristics that are efficient and simpler to implement in the real world.

6.2 Mathematical Formulation

Consider a manufacturing system with a single machine capable of producing two kinds of parts. The system should satisfy a constant demand rate for each part type. The machine is unreliable and is subject to random breakdowns and repairs. We assume that failures and repairs are time dependent (see Gershwin, 1993). The time to failure and time to repair are modeled as exponentially distributed random variables.

The machine is not perfectly flexible in the sense that changeover or setup times between production of different part types are significant. That is,

the machine incurs a nonzero setup time when switching from one product to the other. The setup times are assumed random with exponential distributions. Indeed, In many manufacturing systems, a setup change requires the calibration of the machine and some parts that fail inspection may be initially produced. Thus, requiring a random amount of time. Also, the assumption of a random setup time is very convenient mathematically as will be shown later.

The machine could be either working or under repair. If the machine breaks down, it is immediately brought to repair. During a repair the machine can neither produce parts nor change setups. The machine may breakdown during a setup. This assumption is more realistic than that of no failure of the machine during a setup (see Hu et al., 1995), especially that we consider random setup times. Usually, setups are assumed resumable. That is, in the case of an unreliable system, after a repair, the setup resumes from the point it was left in, just before the failure occurred. This assumption is not always true, especially if the repair involves disassembling parts of the machine or when the manufactured products are chemical fluids for example. In this dissertation, we assume that the setups are nonresumable.

6.2.1 System dynamics

The system considered has a hybrid state comprising both a continuous and a discrete component.

The continuous state equations

Let $x_i(t)$ be the cumulative production surplus of Part Type i (i=1,2) at time t. A positive value of $x_i(t)$ represents inventory, while a negative value represents backlog.

Let $u_i(t)$ be the controlled production rate of the machine producing Type i parts at time t. Let d_i $(i=1,2)$ be a given demand rate for Part Type i.

The rate of change of the surplus of Part Type i is given by

$$\frac{dx(t)}{dt} = u(t) - d, \tag{6.1}$$

where $x(t) = (x_1(t), x_2(t))$, $u(t) = (u_1(t), u_2(t))$, and $d = (d_1, d_2)$.

The discrete event stochastic process

The discrete part of the state variable describes the system's operational mode.

Let $M = \{0,1,2,3,4,5\}$. The operational mode of the machine is described by the random variable $\xi(t)$ with values in M, which is thus the value at time t of a controlled jump process with six possible states. The meaning of each possible state is given in Table 6.1.

Let λ_{ij} be the transition rate from state $i \in M$ to state $j \in M$ at time t.

$$\lambda_{ij} = \lim_{dt \to 0} \frac{1}{dt} \Pr\{\xi(t+dt) = j \,|\, \xi(t) = i\}, \text{ for } (i,j) \in M \times M \tag{6.2}$$

Let Q be the infinitesimal generator of the stochastic process $\{\xi(t), \ t \geq 0\}$. Q is then given by

$$
\begin{pmatrix}
-r & r & 0 & 0 & 0 & 0 \\
p & -p - \lambda_{14} - \lambda_{15} & 0 & 0 & \lambda_{14} & \lambda_{15} \\
p & 0 & -p - \lambda_{25} & 0 & 0 & \lambda_{25} \\
p & 0 & 0 & -p - \lambda_{34} & \lambda_{34} & 0 \\
p & 0 & s_1 & 0 & -p - s_1 & 0 \\
p & 0 & 0 & s_2 & 0 & -p - s_2
\end{pmatrix}
\tag{6.3}
$$

Table 6.1: Operational mode description

$\xi(t)$	Meaning
0	Machine under repair
1	Machine up but not set up
2	Machine changing setup to Part Type 1
3	Machine changing setup to Part Type 2
4	Machine set up for Part Type 1
5	Machine set up for Part Type 2

where p and r are the failure and the repair rates of the machine, respectively. s_1 and s_2 are the rates of changing setup from Part Type 1 to Part Type 2 and from Part Type 2 to Part Type 1, respectively. p, r, s_1, and s_2 are given constants. The transition rates λ_{34}, λ_{25}, λ_{14}, and λ_{15} are all control variables. These are modeled as impulsive controls which trigger an immediate jump of the controlled process (see Haurie and L'Ecuyer (1986)). That is, they can take values only on the set $\{0, \infty\}$. We model these controls as actions through which we decide whether to initiate a new setup for a part type or to continue with the current setup. Let A be the set of actions that the controller can choose from, which is defined by $A = \{a_0, a_1, a_2\}$. These actions are described in Table 6.2.

A state dependent constraint on the action set is defined by a subset Γ of $M \times A$, called the set of admissible state-action couples, such that

$$A(\xi(t)) = \{a(t) \in A \mid (\xi(t), a(t)) \in \Gamma\} \quad \forall \; \xi(t) \in M. \tag{6.4}$$

Table 6.2: Description of the controller action

Action	Meaning
a_0	Do nothing
a_1	Initiate a setup change for Part Type 1
a_2	Initiate a setup change for Part Type 2

Analytically, the set Γ is given by

$$\Gamma = \left\{(0,a_0),(1,a_0),(1,a_1),(1,a_2),(2,a_0),(2,a_2),(3,a_0),(3,a_1),(4,a_0),(5,a_0)\right\} \tag{6.5}$$

The complete state of the system is given by the vector $(x(t),\xi(t)) \in \Re^2 \times M$. The vector $v(t) = (u(t),a(t))$ defines the complete control of the system. Where, $u(t) = (u_1(t),u_2(t))$ specifies the production rates of the two part types at time t, and $a(t)$ specifies the setup decision at time t.

6.2.2 Control constraints

Let U_i denote the maximum production rate of the machine when it is producing Type i parts. Then, we have the following two constraints:

$$0 \le u_1 \le U_1 I_{\{\xi(t)=2\}}; \tag{6.6a}$$

$$0 \le u_2 \le U_2 I_{\{\xi(t)=3\}}, \tag{6.6b}$$

where

$$I_{\{\xi(t)=i\}} = \begin{cases} 1, & \text{if } \xi(t) = i, \ i \in M \\ 0, & \text{otherwise.} \end{cases}$$

The complete control space is defined by the set

$$\Xi(\xi(t)) = \left\{(u(t),a(t)) \mid 0 \le u_i \le U_i I_{\{\xi(t)=(i+1)\}}, \ i = 1, \ 2; \ \text{and } a(t) \in A(\xi(t))\right\}, \tag{6.7}$$

which is a compact set.

6.2.3 Penalty function

We assume that a cost rate of χ_i dollars per unit time ($i=1,2$) is incurred when the machine is undergoing a setup from Part Type j to Part Type i ($j{\neq}i$). The instantaneous cost which penalizes the production for being ahead of (i.e., $x_i > 0$) or being behind (i.e., $x_i < 0$) the demand is given by $\sum_{i=1}^{i=2}(c_i^+ x_i^+ + c_i^- x_i^-)$.

Where, c_i^+ and c_i^- are the per unit instantaneous inventory holding and backlog costs respectively, and $x_i^+(t) = \max\{x_i(t), 0\}$ and $x_i^-(t) = \max\{-x_i(t), 0\}$. The total instantaneous cost is then given by

$$g(x, \xi, \upsilon) = \sum_{i=1}^{i=2} \left(c_i^+ x_i^+ + c_i^- x_i^-\right) + \chi_1 I_{\{\xi=2\}} + \chi_2 I_{\{\xi=3\}}. \tag{6.8}$$

6.2.4 The stochastic control problem

Admissible policies

Let $\{\Im_t : t \geq 0\}$ be an increasing class of sigma-fields describing the history of the process $\{(x(t), \xi(t)) : t \geq 0\}$, and let $(\Omega, \Im)$ be a measurable space. Then, a sample realization $\omega \in \Omega$ corresponds to a trajectory of $\{x(t) : t \geq 0\}$ which is continuous and to a sequence of values of $\{\xi(t) : t \geq 0\}$. The set Π of admissible control policies is a family of $\Im_t$ – adapted processes with values in $\Xi(\xi)$, such that the mapping

$$\pi : \mathfrak{R}^2 \times M \rightarrow \Xi(\xi)$$
$$\pi(x, \xi) \mapsto \upsilon = (u, a) \tag{6.9}$$

is piecewise continuously differentiable (e.g., sufficiently smooth). Thus, with each control policy, $\pi \in \Pi$, is associated a probability measure P_π on $(\Omega, \Im)$ such that the process $\{(x(t), \xi(t)) : t \geq 0\}$ is well defined.

An admissible control policy is a feedback control that specifies the control actions when the system is in a given state. The feedback control determines the production levels and the setup state of the machine as a function of the surplus level and the status of the machine.

Cost functional

The objective is to find a control policy $\pi^* \in \Pi$, corresponding to a production flow rate control $u^* = (u_1^*, u_2^*)$ and a setup control a^*, which

minimizes for each initial condition (x_0, ξ_0) the following expected discounted cost

$$J_\pi(x_0, \xi_0) = E_\pi \left\{ \int_0^\infty e^{-\rho t} g(x(t), \xi(t), v(t)) dt \mid x(0) = x_0, \xi(0) = \xi_0 \right\}, \qquad (6.10)$$

where the minimization is over all functions $\pi(x(\tau), \xi(\tau)) = (u(\tau), a(\tau))$, such that $x(\tau)$, $\xi(\tau)$, $u(\tau)$ and $a(\tau)$ satisfy the system dynamics and $(u(\tau), a(\tau)) \in \Xi(\xi)$ for $\tau \in [0, +\infty)$. Here, ρ is a given positive discount rate.

6.2.5 Optimality necessary and sufficient conditions

This optimal control problem belongs to the class of problems considered by Rishel (1975). It is to control a stochastic process with jump Markov disturbances. A necessary and sufficient condition characterizing the optimal control policy, under appropriate differentiability assumptions of the control (Rishel (1975)), are described by the following set of partial differential equations (Known as the Hamilton-Jacobi-Bellman (HJB) optimality equations):

$$\rho J(x, \xi) = \min_{v \in \Xi(\xi)} \left\{ g(x, \xi, v) + \sum_{i=1}^{i=2} \frac{\partial J(x, \xi)}{\partial x_i} \dot{x}_i + \sum_{\xi' \in M} q_{\xi\xi'}(v) J(x, \xi') \right\}, \quad \forall \xi \in M \qquad (6.11)$$

where $J(x, \xi)$ denotes the optimal cost value. $q_{\xi\xi'}(v)$ is the transition rate from state ξ to state ξ', which depends on the control v. Here, $\dot{x}_i = u_i - d_i$ denotes the total derivative of x_i with respect to time. Notice that the production rates u_i ($i = 1, 2$) should be chosen so as to minimize the rate at which the cost increases with respect to x_i. The setup decisions are determined independently from the u_i's.

Based on Rishel's (1975) formalism, the optimal control formulation has been applied to a number of flexible manufacturing systems (Tsitsiklis (1982), Akella & Kumar (1986), and Kimemia and Gershwin (1983)). Although there is

no available solution technique to solve this control problem analytically, the previous work has revealed some characteristics of the optimal solution. A general finding is that the optimal solution is a switching policy. That is, the control policy is determined by a partition of the state space, and control actions take place when the system state moves from one region to another. In the next section, we adopt a numerical approach to obtain the optimal solution of the control problem. We use an approximation technique due to Kushner and Dupuis (1992) to discretize the HJB partial differential equations. This technique leads to an infinite horizon stochastic dynamic programming algorithm which we solve using the successive approximation technique. Then, based on the numerical solution, we observe the structure of the optimal solution and construct a heuristic for real-time scheduling.

6.3 Numerical Approach

6.3.1 Markov chain approximation method

In this section, we adopt Kushner and Dupuis's (1992) method to construct an approximating discrete time discrete state Markov chain to the original continuous stochastic control problem. we have deliberately tried to keep the notation as close as possible to that used by Kushner and Dupuis.

The general idea of Kushner and Dupuis's approach is to approximate the value function $J(x, \xi)$ by $J^h(x, \xi)$ via a finite difference method. Basically, the partial derivatives $\partial J(x_i, \xi)/\partial x_i$ are approximated by the following one sided difference approximations:

$$\partial J(x_i, \xi)/\partial x_i = \begin{cases} [J(x + h_i e_i, \xi) - J(x, \xi)]/h_i & \text{if } \dot{x}_i \geq 0 \\ [J(x, \xi) - J(x - h_i e_i, \xi)]/h_i & \text{if } \dot{x}_i < 0 \end{cases} \tag{6.12}$$

where $h_i \in \mathfrak{R}^+$ represents the finite difference interval of the variable x_i. $e_i \in \mathfrak{R}^2$ is the unit vector in the ith coordinate direction. Notice here that the direction (forward or backward) of the finite difference approximation (6.12) is the direction of the production surplus velocity component x_i.

Let $b_i(x) = \dot{x}_i(t)$ and define the following quantities: $b_i^+(x) = \max(b_i(x), 0)$ and $b_i^-(x) = \max(-b_i(x), 0)$. Then,

$$\frac{\partial J^h(x,\xi)}{\partial x_i} \dot{x}_i = \frac{J^h(x + h_i e_i, \xi) - J^h(x,\xi)}{h_i} b_i^+(x)$$

$$+ \frac{J^h(x,\xi) - J^h(x - h_i e_i, \xi)}{h_i} b_i^-(x). \qquad (6.13)$$

Substituting in (6.11) and using the fact that $q_{aa}(v) = -\sum_{a' \neq a} q_{aa'}(v)$, we obtain

$$\left(\rho + \sum_{\xi' \neq \xi} q_{\xi\xi'}(v) \right) J^h(x,\xi) = \min_{v \in \Pi(\xi)} \left\{ g(x,\xi,v) - \sum_{i=1}^{i=2} \frac{\left(b_i^+(x) + b_i^-(x) \right)}{h_i} J^h(x,\xi) \right.$$

$$\left. + \sum_{i=1}^{i=2} \left[\frac{b_i^+(x)}{h_i} J^h(x + h_i e_i, \xi) + \frac{b_i^-(x)}{h_i} J^h(x - h_i e_i, \xi) \right] + \sum_{\xi' \neq \xi} q_{\xi\xi'}(v) J^h(x, \xi') \right\};$$

Let $Q^h(x, \xi, v) = \sum_{\xi' \neq \xi} q_{\xi\xi'}(v) + \sum_{i=1}^{i=2} \left(|b_i(x)| / h_i \right)$. Noticing that $b_i^+(x) + b_i^-(x) = |b_i(x)|$, dividing through by $Q^h(x, \xi, v)$ in the previous equation and rearranging terms, gives

$$\left(1 + \frac{\rho}{Q^h(x, \xi, v)} \right) J^h(x, \xi) = \min_{v \in \Pi(\xi)} \left\{ \frac{g(x, \xi, v)}{Q^h(x, \xi, v)} \right.$$

$$\left. + \sum_{i=1}^{i=2} \left[\frac{b_i^+(x)}{Q^h(x, \xi, v) h_i} J^h(x + h_i e_i, \xi) + \frac{b_i^-(x)}{Q^h(x, \xi, v) h_i} J^h(x - h_i e_i, \xi) \right] + \sum_{\xi' \neq \xi} \frac{q_{\xi\xi'}(v)}{Q^h(x, \xi, v)} J^h(x, \xi') \right\};$$

Define the grid S^h where each component of the surplus vector x_i can only take values that are positive or negative multiples of h_i. That is,

$$S^h = \{x \mid x = \sum_i h_i e_i n_i \ : \ n_i = 0, \pm 1, \pm 2, \ldots, \ i = 1, 2\}.$$

Let $p^h(x, \xi, \xi', v) = \dfrac{q_{\xi\xi'}(v)}{Q^h(x, \xi, v)}$, and let

$$p^h(x, x', \xi, v) = \begin{cases} \dfrac{b_i^{\pm}(x)}{Q^h(x, \xi, v) h_i} & \text{if } x' = x \pm h_i e_i \\ 0 & \text{for all other} x' \in S^h \end{cases}$$

Then, we obtain

$$J^h(x, \xi) = \min_{v \in U(\xi)} \left\{ \frac{g(x, \xi, v)}{Q^h(x, \xi, v)\left(1 + \dfrac{\rho}{Q^h(x, \xi, v)}\right)} \right.$$

$$\left. + \frac{1}{\left(1 + \dfrac{\rho}{Q^h(x, \xi, v)}\right)} \left[\sum_{x' \in S^h} p^h(x, x', \xi, v) J^h(x', \xi) + \sum_{\xi' \neq \xi} p^h(x, \xi, \xi', v) J^h(x, \xi) \right] \right\}. \quad (6.14)$$

Equations (6.14) have a nice interpretation. Indeed, they represent the dynamic programming equations of a discounted cost infinite horizon discrete time discrete state Markov Decision Process. Where, $p^h(x, x', \xi, v)$ and $p^h(x, \xi, \xi', v)$ are the transition probabilities (since they are nonnegative and sum to 1 over $(x', \xi') \in \prod_{\xi' \in M} S^h$ for each $(x, \xi) \in \prod_{\xi \in M} S^h$, and $1/(1 + \rho/Q^h(x, \xi, v)) < 1$ is the discount factor.

In Kushner and Dupuis (1992), it is shown that this approximating Markov chain satisfies the contraction properties which guarantee the existence of a solution which can be obtained by classical iterative methods in policy or value space. Also, it is shown that with appropriate boundary conditions $\lim_{h \to 0} J^h(x, \xi) = J(x, \xi)$.

6.3.2 Computational algorithm: The value iteration method

For any function $J^h : \prod_{\xi \in M} S^h \to \Re$, $\pi : \prod_{\xi \in M} S^h \to \Xi(\xi)$ with $\pi \in \Pi$, for all $(x, \xi) \in \prod_{\xi \in M} S^h$, we denote:

$$
T_\pi(J^h)(x, \xi) = \frac{1}{\left(1 + \dfrac{\rho}{Q^h(x, \xi, \pi(x, \xi))}\right)} \left\{ \frac{g(x, \xi, \pi(x, \xi))}{Q^h(x, \xi, \pi(x, \xi))} \right.
$$

$$
\left. + \sum_{x' \in S^h} p^h(x, x', \xi, \pi(x, \xi)) J^h(x', \xi) + \sum_{\xi' \neq \xi} p^h(x, \xi, \xi', \pi(x, \xi)) J^h(x, \xi) \right\}; \qquad (6.15)
$$

Note that $T_\pi(J^h)(.,.)$ is a function defined on $\prod_{\xi \in M} S^h$, and T_π may be viewed as a mapping that transforms a function J^h on $\prod_{\xi \in M} S^h$ into another function $T_\pi(J^h)$ on $\prod_{\xi \in M} S^h$. The mapping T_π plays an important theoretical role and provides a convenient shorthand notation in expressions that would be too complicated to write otherwise. from (6.15) it can be seen that $T_\pi(J^h)(x, \xi)$ represents the optimal cost corresponding to policy $\{\pi, \pi, ...\}$ for the one-stage problem with initial state (x, ξ), stage cost $g(x, \xi, \pi(x, \xi)) / (\rho + Q^h(x, \xi, \pi(x, \xi)))$, and terminal cost function

$$
Q^h(x, \xi, \pi(x, \xi)) \left[\sum_{x' \in S^h} p^h(x, x', \xi, \pi(x, \xi)) J^h(x', \xi) \right.
$$

$$
\left. + \sum_{\xi' \neq \xi} p^h(x, \xi, \xi', \pi(x, \xi)) J^h(x, \xi) \right] / (\rho + Q^h(x, \xi, \pi(x, \xi))).
$$

Let $T^*(J^h)(x, \xi) = \min_{\pi \in \Pi} \{ T_\pi(J^h)(x, \xi) \}$

Algorithm

Given a finite difference $h \in \Re_+^2$:

Initialization step:

Choose $\varepsilon \in \Re_+$;

Let $k := 1$, and $\left(J^h\right)^k(x, \xi) = g\left(x, 1, (0, a_0)\right) / \left(1 - e^{-\rho}\right)$, $\forall (x, \xi) \in \prod_{\xi \in M} S^h$;

Go to main step.

Main step:

Let $\left(J^h\right)^{k-1}(x, \xi) := \left(J^h\right)^k(x, \xi)$, $\forall (x, \xi) \in \prod_{\xi \in M} S^h$;

determine the policy π^k such that:

$\left(J^h\right)^k(x, \xi) := T_{\pi^k}\left(J^h\right)^{k-1}(x, \xi) := T^*\left(J^h\right)^{k-1}(x, \xi)$, $\forall (x, \xi) \in \prod_{\xi \in M} S^h$;

Let $\underline{c_k} := \min_{\forall (x, \xi) \in \prod_{\xi \in M} S^h} \frac{\rho}{1 - \rho}\left\{\left(J^h\right)^k(x, \xi) - \left(J^h\right)^{k-1}(x, \xi)\right\}$;

$\overline{c_k} := \max_{\forall (x, \xi) \in \prod_{\xi \in M} S^h} \frac{\rho}{1 - \rho}\left\{\left(J^h\right)^k(x, \xi) - \left(J^h\right)^{k-1}(x, \xi)\right\}$.

Go to Convergence test step.

Convergence test step:

If $\left|\overline{c_k} - \underline{c_k}\right| > \varepsilon$, Let $k := k + 1$ and go to main step;

else $\pi^* = \pi^k$, stop.

6.3.3 Implementation considerations

Since Grid S^h is finite, we have to impose certain boundary conditions on the states of the system that fall outside Grid S^h. These boundary conditions constitute an additional approximation to the original problem. However, as will be shown from the subsequent numerical examples, they do not affect the optimal solution. For each $\xi \in M$, whenever a component x_i of x is on the boundary, the boundary conditions consist of approximating the cost of the state to which the system is going to jump, which is outside Grid S^h, by a linear interpolation between the current state and another state which is inside Grid S^h. To better understand these boundary conditions, consider the following situation: Assume that x_i is on the boundary of S^h and that $x_i + h_i$ is

outside S^h. Hence, $x_1 - h_1$ is inside S^h. A linear interpolation, using $J^h(x_1, x_2, \xi)$ and $J^h(x_1 - h_1, x_2, \xi)$, gives the following approximation for $J^h(x_1 + h_1, x_2, \xi)$:

$$J^h(x_1 + h_1, x_2, \xi) = 2J^h(x_1, x_2, \xi) - J^h(x_1 - h_1, x_2, \xi).$$

In the following section, we illustrate this numerical approach by means of two examples.

6.4 Optimal Policy

To characterize the optimal policy, we apply the above algorithm to two cases which data are given in Table 6.3.

For both cases, the discount rate $\rho = 0.1$. The grid S^h is defined as follows: $S^h = \{x | x = \sum_i h_i e_i n_i : h_i = 0.5, \ n_i = -30,\ldots,0,\ldots20, \ i = 1,2\}$.

Table 6.3: Data for cases 1 and 2.

	Case 1	Case 2
p	0.05	0.05
r	0.9	0.9
(d_1, d_2)	(0.25, 0.25)	(0.25, 0.25)
(U_1, U_2)	(0.67, 0.67)	(0.67, 0.67)
(s_1, s_2)	(0.5, 0.5)	(0.5, 0.5)
(c_1^+, c_1^-)	(1.0, 10.0)	(1.5, 15.0)
(c_2^+, c_2^-)	(1.0, 10.0)	(1.0, 10.0)
(χ_1, χ_2)	(1.0, 1.0)	(2.0, 1.0)

The results are shown in Figures 6.1.a, 6.1.b, 6.1.c, 6.1.d and 6.3 for Case 1 and Figures 6.2.a, 6.2.b, 6.2.c, 6.2.d and 6.4 for Case 2.

6.4.1 Optimal setup policy when the machine is up

As shown in Figures 6.1.a and 6.2.a, the optimal setup policy partition the state space (for $\xi = 2$ or $\xi = 3$) into four mutually exclusive regions. These are described below.

Region I: If the machine is already set up for the production of Part Type 1, then it is optimal to continue with this setup and produce this part type at the maximum production rate. On the other hand, If the machine is set up for the production of Part Type 2, then it is optimal to setup for Part Type 1 immediately and start producing this part type at maximum production rate.

Notice the difference in the shape of Region I in Cases 1 and 2. This is mainly due to the fact that Part Type 1 needs to be produced more often than Part Type 2, since its backlog, inventory, and setup costs are higher than those of Part Type 2.

Region II: The policy in this region is just the opposite of that of Region I. That is, a setup is optimal only if the machine is already set up for the production of Part Type 1. The production strategy is to produce at maximum production rate when the machine is set up for the production of the proper part type.

Region III: The policy in this region is to keep the same setup of the machine and set the production rate to zero. In other words the production is stopped for either part type.

Region IV: If the system is in this region, the production continues with the current setup at maximum rate. Regions I, II, and III are transient and Region IV is recurrent (i.e., the surplus vector (x_1, x_2) will be trapped with probability one in Region IV, once the system moves into that region). To see this, consider Figure 6.5, where different possible trajectories of the production

surplus vector are shown to endup in Region IV no matter where the initial production surplus originated. Trajectory 1 (thick solid line) shows a surplus starting in Region II, with the machine set up for Part Type 2. The policy of Region II dictates that we should start a setup immediately and produce Part Type 1 at the maximum production rate. This will drive the surplus trajectory to Region IV. Trajectories 3 and 4 (dashed and dotted lines respectively) show a starting surplus in Region III, in which the policy dictates that we stop production (surplus levels too high). The surplus of both parts decreases and either boundary of Region II (trajectory 3) or boundary of Region I (trajectory 4) is encountered. At this point, a chattering on the boundary occurs, since we produce and stop the machine alternately. This chattering continuous until the surplus vector enters Region IV, where it stays thereafter. Trajectory 2 has the same behavior as Trajectory 1.

6.4.2 Optimal setup policy right after repair

As shown in Figures 6.1.b and 6.2.b it is optimal to start a setup for the production of Part Type 1 (Part Type 2) if the production surplus vector is in Region I (Region II). If the production surplus vector is in Region III, then it is better to delay the setup until the production surplus vector, driven by the demand, drifts to either Region I or II.
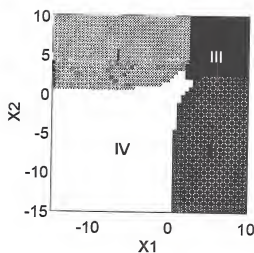
Figure 6.1.a: Optimal setup policy
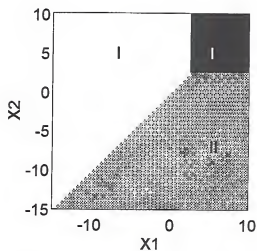(machine is up)

Figure 6.1.b: Optimal setup policy
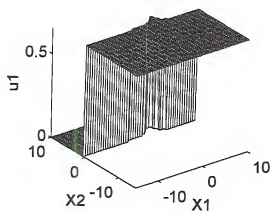right after repair

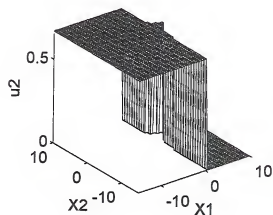Figure 6.1.c: Optimal production
rate for Part Type 1

Figure 6.1.d: Optimal production
rate for Part Type 2

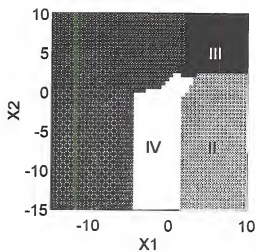Figure 6.1: Optimal policy for Case 1

Figure 6.2.a: Optimal setup policy
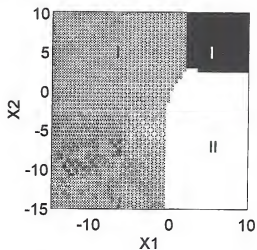(machine is up)

Figure 6.2.b: Optimal setup policy
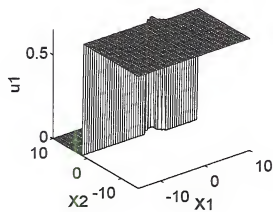right after repair

Figure 6.2.c: Optimal production
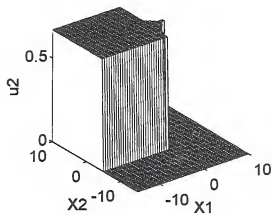rate for Part Type 1

Figure 6.2.d: Optimal production
rate for Part Type 2

Figure 6.2: Optimal policy for Case 2

6.4.3 Optimal production policy

Figures 6.1.c, 6.1.d, 6.2.c and 6.2.d show the optimal production rates of the machine (when it is possible to produce) for Cases 1 and 2. As can be seen, the optimal policy is of the bang-bang control type. That is, the production rate is either at its lower (zero in this case) or at its maximum level, and there is no optimal production level in between.

6.4.4 Optimal cost function

In Figures 6.3 and 6.4, we give, for Case 1 and Case 2 respectively, the optimal cost as a function of the surplus vector, as well as its corresponding contour lines for each setup (i.e., when the machine is set up to produce Part Type 1 and when it is set up to produce Part Type 2). Notice that the optimal cost function is convex (the level sets are convex). Also, it is very important to notice that the cost function is everywhere continuously differentiable which is the basic assumption for the HJB partial differential equations.

6.4.5 Policy structure

In both cases the optimal policy is of the corridor type as proposed by Sharifnia et al. (1991) (Region IV is the corridor). The sample trajectory shown in Figure 6.6 shows that the production surplus vector will converge to an average cyclic schedule. This cyclic schedule was referred to as the *Limit Cycle* in Sharifnia et al. (1991). In the following, we formally define the two types of optimal policy structures obtained.

Definition 6.1: We call an Orthogonal Corridor Policy (OCP), a policy having the setup policy shown in Figure 6.1.a (Case 1).
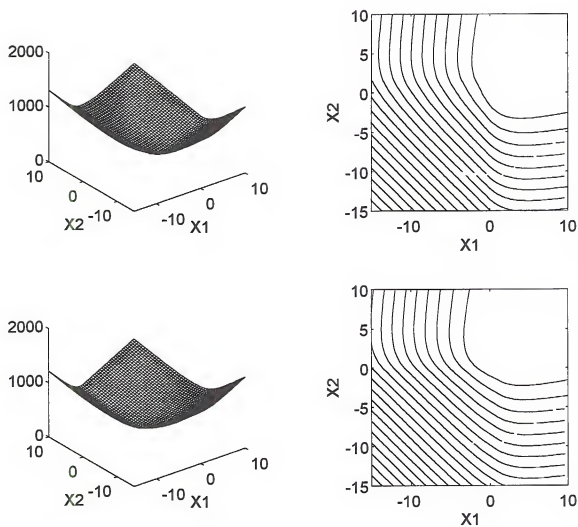
Figure 6.3 : $J(x_1,x_2,2)$ (upper 2 graphs) and $J(x_1,x_2,3)$ (lower 2 graphs) for Case 1.

Figure 6.4 : $J(x_1, x_2, 2)$ (upper 2 graphs) and $J(x_1, x_2, 3)$ (lower 2 graphs) for Case 2.

Definition 6.2: We call a Parallel Corridor Policy (PCP), a policy having the setup policy shown in Figure 6.2.a (Case 2).

In the next section, we give a necessary and sufficient condition for the demand to be feasible. Then, we propose a heuristic based on the results of the deterministic system studied in Chapter 5 to approximate the optimal policy of sufficiently reliable stochastic systems.

Figure 6.5: Production surplus trajectories emanating in different regions.

Figure 6.6: Typical surplus trajectory within the corridor.

6.5 Demand Feasibility

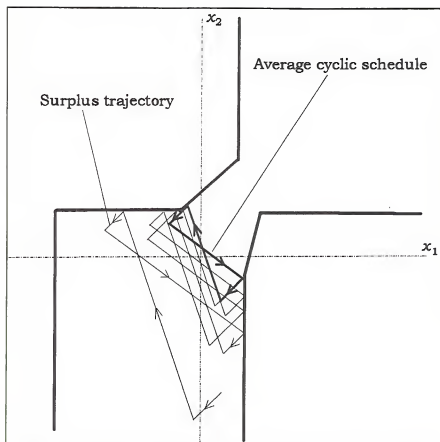In this section, we provide a necessary and sufficient condition for the demand to be feasible. Applying the conditions stated below allows us to distinguish between systems with enough capacity to meet the demand and those for which the demand will never be met. In the latter case, no matter which policy we use, the cumulative production surplus of the system will eventually drift to $-\infty$.

The following proposition gives a necessary and sufficient condition for the system to meet the demand.

Proposition 6.1 (*Feasibility Test*): The demand set, $d = \{d_1, d_2\}$, is feasible if and only if

$$U_i\left(\frac{r}{p+r}\right)\left(\frac{s_i}{p+s_i}\right) \geq d_1 + d_2, \qquad (i=1,2)$$

Proof: The proof is based on a setup policy referred to as the Maximum Capacity Policy (MCP). Let T be the planning horizon and without loss of generality, we assume that the machine is operational and has not been set up at time $t = 0$. The MCP operates as follows: Start with a setup for Part Type 1 at time $t = 0$, then keep producing this type at maximum production rate U_1, until time $t = T_1$. During the time interval $(0, T_1)$, the machine is set up for Part Type 1 after each repair (since the setup is nonresumable). At time $t = T_1$, we switch to Part Type 2 and continue producing this type at maximum production rate U_2, until time $T = T_1 + T_2$. During the time interval (T_1, T), the machine is set up for Part Type 2 after each repair. To maximize the production capacity of the system, under such a control policy, we choose T_1 and T_2 so as $\frac{T_1}{T_2} = \frac{d_1}{d_2}$.

Therefore, we have $T_1 = \frac{d_1}{d_1 + d_2}T$ and $T_2 = \frac{d_2}{d_1 + d_2}T$.

Note that, operating the system in this fashion, the number of setups is minimum and the machine is always producing at its maximum production rate when it is up. Therefore, it allows the maximum attainable system capacity. We want to show that under MCP, the condition above is necessary and sufficient for the demand to be feasible.

Now, let $N(t)$, $t > 0$ be the number of machine failures at time t, i.e., $\{N(t), t > 0\}$ is a renewal process with inter-renewal times $t_i = t_i^u + t_i^d$ ($i = 1, 2, \ldots$). Where t_i^u, t_i^d ($i = 1, 2, \ldots$) are exponentially distributed with rates p and r respectively. Also, let $P_i(T)$ ($i=1,2$) be the cumulative production at time T of Part Type i. Then we have

$$P_i(T) = \left(\sum_{i=0}^{N(T_i)} t_i^u - N(T_i)\Delta_i \right) U_i, \quad (i=1,2).$$

Let Δ_i be the random variable denoting the total time spent to complete a setup. Δ_i includes the random setup time and the time spent in repair when a breakdown occurs during a setup.

The demand is feasible if and only if the following holds:

$$\lim_{T \to \infty} \frac{E[P_i(T)]}{T} \geq d_i, \quad (i=1,2).$$

$$\Leftrightarrow \quad \lim_{T \to \infty} \frac{E\left[\sum_{i=0}^{N(T_i)} t_i^u \right] - E[N(T_i)]E[\Delta_i]}{T} \geq \frac{d_i}{U_i}, \quad (i=1,2).$$

$$\Leftrightarrow \quad \lim_{T \to \infty} \frac{E[N(T_i)]}{T} \left(E[t_i^u] - E[\Delta_i] \right) \geq \frac{d_i}{U_i}, \quad (i=1,2).$$

$$E[t_i^u] = \frac{1}{p} \text{ and } E[\Delta_i] = \frac{1}{p+s_i}.$$

$$\Leftrightarrow \quad \lim_{T \to \infty} \frac{E[N(T_i)]}{T_i} \left(\frac{T_i}{T} \right) \frac{s_i}{p(p+s_i)} \geq \frac{d_i}{U_i}, \quad (i=1,2).$$

$$\Leftrightarrow \quad \lim_{T \to \infty} \frac{E[N(T_i)]}{T_i} \left(\frac{d_i}{d_1+d_2} \right) \frac{s_i}{p(p+s_i)} \geq \frac{d_i}{U_i}, \quad (i=1,2).$$

Note that, when $T \to \infty$, $T_i \to \infty$. Using the elementary renewal Theorem of Renewal Theory, we have $\lim_{T_i \to \infty} \frac{E[N(T_i)]}{T_i} = \frac{1}{\mu}$. Where $\mu = \frac{1}{p} + \frac{1}{r}$. The result follows immediately. ■

Remark: Intuitively, if the condition holds, then we know that there is at least one setup policy (the MCP precisely) for which the demand is feasible. On the other hand, if Proposition 6.1 is violated then we know for sure that there is

no other policy for which the demand will be feasible, since the MCP which allows the maximum attainable capacity, failed to satisfy the feasibility test.

6.6 Near Optimal Real-Time Policy for Sufficiently Reliable Systems

In this section, we propose a heuristic, implementable in real-time, for stochastic systems that are sufficiently reliable. The heuristic is based on the results of the deterministic system studied in the previous chapters. it is referred to as the hedging corridor policy which is characterized by the two corridor edges defined by z_1 and z_2 as shown in Figure 6.7. Notice that, before we switch to the production of the other part type, a certain inventory is built to hedge against future shortages brought about machine failures and setups.

6.6.1 Construction of the hedging corridor policy

The hedging corridor policy is constructed as follows: We notice, from the numerical optimal solution, that all trajectories converge to a certain average cyclic schedule and that there is no production at the demand rate. Hence, there must be no windows in the corridor (since the trajectories will never reach the cyclic schedule in finite time), and the constructed corridor should be positioned on the extracted schedule rather than on the cyclic schedule itself (see Chapter 5). Hence, the hedging corridor policy is completely characterized by

$$z_1 = A_1 \text{ and } z_2 = C_2,$$

where A_1 is the abscissa of Point A, and C_2 is the ordinate of Point C defined in Chapter 5. To be able to use the formulas of A_1 and C_2, we substitute the setup times δ_1 and δ_2 by the average setup times of the stochastic system, given by

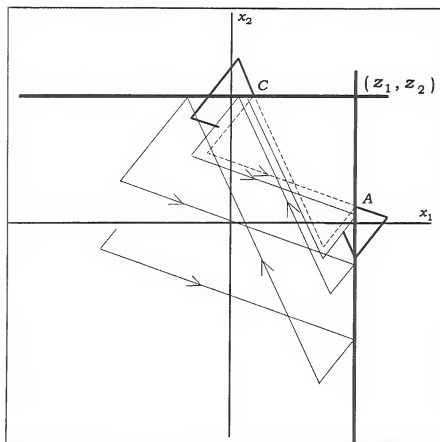$$\delta_1 = \frac{1}{s_1} \text{ and } \delta_2 = \frac{1}{s_2}.$$

Figure 6.7: Typical surplus trajectory within the hedging corridor.

6.6.2 Performance of the hedging corridor policy

Study cases

To test the performance of the hedging corridor policy, we simulated the system using the optimal numerical policies and then compared with the results using the heuristic. We used the cases shown in Table 6.4. Case 1 is referred to as the symmetric case, since both part types have the same characteristics. Compared to the symmetric case:

Cases 2 and 3 show an in increase of the setup time of Part Type 1.

Cases 4, 5, and 6 show an increase of the inventory cost rate of Part Type 1.

Cases 7, 8, and 9 show an increase of the backlog cost rate of Part Type 1.

Cases 10 and 11 show an increase of the setup cost rate of Part Type 1.

Cases 12, 13, and 14 show an increase of the production rate of Part Type 1.

Cases 15, 16, and 17 show an increase of the demand rate of Part Type 1.

For Cases 1 through 17, we used the following data: $\rho = 0.1$, $p = 0.05$, $r = 0.9$ and a grid $S^h = \{x | x = \sum_i h_i e_i n_i : h_i = 0.5, \ n_i = -30,...,0,...20, \ i = 1,2\}$.

To test the sensitivity of the hedging corridor policy to the failure rate of the machine, we increased p using the symmetric case (Case 1).

$p = 0.1$ corresponds to Case 18.

$p = 0.15$ corresponds to Case 19.

$p = 0.2$ corresponds to Case 20.

<u>Performance measures</u>

- The *total discounted cost* of operating the system over the simulation horizon.

- The *average surplus level* $\overline{X}_i$ for Part Type i over the simulation horizon. The closer the average surplus to zero, the better the policy.

- The *production percentage,* given as follows:

$$P_i = \frac{W_i}{D_i} \times 100\%, \quad \text{for } i = 1,2,$$

where W_i and D_i are the cumulative production and cumulative demand of Part Type i over the simulation horizon. This is the fraction of the satisfied demand of Part Type i expressed in percentage. The closer this measure to 100%, the better the policy.

- The *aggregate production percentage* given as follows:

$$P = \frac{\sum_{i=1}^{2} W_i}{\sum_{i=1}^{2} D_i} \times 100\%.$$

- The *total average surplus* $\overline{X}$, which is the sum of the individual part type average surpluses.

Table 6.4: Simulation cases

Cases	c_1^-	c_1^+	c_2^-	c_2^+	χ_1	χ_2	U_1	U_2	d_1	d_2	s_1	s_2
1	15	1.5	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	1	1
2	15	1.5	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	0.75	1
3	15	1.5	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	0.5	1
4	15	2	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	1	1
5	15	2.5	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	1	1
6	15	3	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	1	1
7	20	1.5	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	1	1
8	25	1.5	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	1	1
9	30	1.5	15	1.5	0.5	0.5	1.13	1.13	0.32	0.32	1	1
10	15	1.5	15	1.5	0.75	0.5	1.13	1.13	0.32	0.32	1	1
11	15	1.5	15	1.5	1	0.5	1.13	1.13	0.32	0.32	1	1
12	15	1.5	15	1.5	0.5	0.5	1.43	1.13	0.32	0.32	1	1
13	15	1.5	15	1.5	0.5	0.5	1.83	1.13	0.32	0.32	1	1
14	15	1.5	15	1.5	0.5	0.5	2.26	1.13	0.32	0.32	1	1
15	15	1.5	15	1.5	0.5	0.5	1.13	1.13	0.4	0.32	1	1
16	15	1.5	15	1.5	0.5	0.5	1.13	1.13	0.48	0.32	1	1
17	15	1.5	15	1.5	0.5	0.5	1.13	1.13	0.64	0.32	1	1

<u>Simulation</u>

A continuous simulation model was implemented. For each of the 20 cases above, the simulation was replicated 10 times using different random variate streams (to reduce biases). The simulation horizon (for one replication) was fixed to 10,000 time units. The initial surplus values were equal to zero in all the cases.

Simulation results

Tables 6.5 and 6.6 show the simulation results using the optimal policies and the hedging corridor policy, respectively. Besides the performance measures defined above, Tables 6.5 and 6.6 show the structure of the optimal policy and the result of the feasibility test for each case.

Table 6.6 shows that the hedging corridor policy performs very well compared to the optimal policy. Notice that for all the cases the demand is met at 100% and that the average production surplus levels are practically 0. The maximum percentage gap between the total discounted cost of the optimal policies and the heuristic ones is 23.9% and occurs with Case 17 for which the demand of Part Type 1 is double the demand of Part type 2.

We might conclude that for sufficiently reliable systems producing part types that are similar in characteristics (which is usually the case in practice), the hedging corridor policy performs very well. The latter also performs well with sufficiently reliable and moderately loaded systems. When the system becomes heavily loaded or frequently unavailable the heuristic performs poorly.

Notice that for some cases, the total discounted cost of the heuristic policy outperformed that of the optimal policy. This is mainly due to the continuous implementation of the simulation. Indeed, when simulating the system using the hedging corridor policy, we calculate exactly the quantity of a part type to be produced before we switch to the other. While, when simulating the system using the optimal policy, we calculate the quantity to be produced during one time unit and then check whether we hit the switching boundary or not. This might cause an overshooting of the boundary resulting in higher cost (especially, when the production rates are high compared to the demand rates).

Table 6.5: Simulation results using the numerical optimal policy

Cases	Demand Feasibility	Policy Structure	Total cost	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}$	P_1	P_2	P
1	Feasible	OCP	100.7	0.0	0.0	0.0	100.0	100.0	100.0
2	Feasible	OCP	114.3	0.0	0.0	0.0	100.0	100.0	100.0
3	Feasible	OCP	149.9	0.0	0.0	0.0	100.0	100.0	100.0
4	Feasible	OCP	103.2	0.0	0.0	0.0	100.0	100.0	100.0
5	Feasible	OCP	103.6	0.0	0.0	0.0	100.0	100.0	100.0
6	Feasible	OCP	105.1	0.0	0.0	0.0	100.0	100.0	100.0
7	Feasible	PCP	107.4	0.0	0.0	0.0	100.0	100.0	100.0
8	Feasible	PCP	111.8	0.0	0.0	0.0	100.0	100.0	100.0
9	Feasible	PCP	112.0	0.0	0.0	0.0	100.0	100.0	100.0
10	Feasible	OCP	101.0	0.0	0.0	0.0	100.0	100.0	100.0
11	Feasible	OCP	101.3	0.0	0.0	0.0	100.0	100.0	100.0
12	Feasible	PCP	92.9	0.0	0.0	0.0	100.0	100.0	100.0
13	Feasible	PCP	85.0	0.0	0.0	0.0	100.0	100.0	100.0
14	Feasible	PCP	82.0	0.0	0.0	0.0	100.0	100.0	100.0
15	Feasible	OCP	137.6	0.0	0.0	0.0	100.0	100.0	100.0
16	Feasible	OCP	170.4	0.0	0.0	0.0	100.0	100.0	100.0
17	Feasible	OCP	291.8	0.0	0.0	0.0	100.0	99.8	99.9
18	Feasible	OCP	146.4	0.0	0.0	0.0	100.0	100.0	100.0
19	Feasible	OCP	178.3	0.0	0.0	0.0	100.0	100.0	100.0
20	Feasible	OCP	207.5	0.0	0.0	0.0	100.0	100.0	100.0

Table 6.6: Simulation results using the hedging corridor policy

Cases	Demand Feasibility	Policy Structure	Total cost	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}$	P_1	P_2	P
1	Feasible	OCP	104.1	0.0	0.0	0.0	100.0	100.0	100.0
2	Feasible	OCP	125.0	0.0	0.0	0.0	100.0	100.0	100.0
3	Feasible	OCP	175.4	0.0	0.0	0.0	100.0	100.0	100.0
4	Feasible	OCP	106.4	0.0	0.0	0.0	100.0	100.0	100.0
5	Feasible	OCP	108.8	0.0	0.0	0.0	100.0	100.0	100.0
6	Feasible	OCP	111.1	0.0	0.0	0.0	100.0	100.0	100.0
7	Feasible	PCP	110.4	0.0	0.0	0.0	100.0	100.0	100.0
8	Feasible	PCP	116.7	0.0	0.0	0.0	100.0	100.0	100.0
9	Feasible	PCP	123.0	0.0	0.0	0.0	100.0	100.0	100.0
10	Feasible	OCP	104.7	0.0	0.0	0.0	100.0	100.0	100.0
11	Feasible	OCP	105.2	0.0	0.0	0.0	100.0	100.0	100.0
12	Feasible	PCP	89.0	0.0	0.0	0.0	100.0	100.0	100.0
13	Feasible	PCP	75.3	0.0	0.0	0.0	100.0	100.0	100.0
14	Feasible	PCP	71.5	0.0	0.0	0.0	100.0	100.0	100.0
15	Feasible	OCP	131.2	0.0	0.0	0.0	100.0	100.0	100.0
16	Feasible	OCP	184.0	0.0	0.0	0.0	100.0	100.0	100.0
17	Feasible	OCP	361.4	0.0	0.0	0.0	100.0	99.8	99.9
18	Feasible	OCP	160.0	0.0	0.0	0.0	100.0	100.0	100.0
19	Feasible	OCP	202.7	0.0	0.0	0.0	100.0	99.9	100.0
20	Feasible	OCP	255.5	0.0	0.0	0.0	100.0	100.0	100.0

6.7 Summary

In this chapter, we studied an unreliable one-machine two-part-type manufacturing system with random nonresumable setups. We formulated the problem as a stochastic optimal control problem. To obtain the optimal setup and production control policy, we used a simple technique due to Kushner and Dupuis (1992) and converted the problem to a classical infinite horizon discrete time discrete state Markov Decision Process. The latter was solved numerically using the value iteration algorithm of dynamic programming. The optimal policy was shown to have two different structures (Orthogonal Corridor Policy or Parallel Corridor Policy). Based on results of the corresponding deterministic system, a heuristic, implementable in real-time, was proposed. Simulation results show that the heuristic policy performs very well when applied to systems that are sufficiently reliable producing similar part types or not heavily loaded.

CHAPTER 7
SUMMARY AND SUGGESTIONS FOR FUTURE RESEARCH

7.1 Summary

In this dissertation, a study of a manufacturing system, comprising a single machine that is able to produce different part types or products, has been presented. The manufacturing system is inflexible in the sense that when production is switched from one part type to the next a setup time and setup cost are incurred. The manufacturing system has a finite production capacity, and has to satisfy a constant demand, for each part type, over a certain planning horizon. Two different scenarios has been considered: A manufacturing system with a perfectly reliable machine, which we referred to as the deterministic system, and a manufacturing system with an unreliable machine subject to random failures, random repairs and random setup times, which we referred to as the stochastic system.

The study of the deterministic system has been presented in Chapters 2, 3, 4, and 5. In Chapter 2, the problem has been formulated as an optimal control problem. In Chapter 3, the optimal steady state solution of a two-part-type system has been obtained analytically. The problem was then extended to the multi-part-type case. For the latter, a nonlinear programming algorithm, to obtain the optimal steady state solution, was proposed. In Chapters 4 and 5, a study of the transient behavior of the deterministic two-part-type system was presented. In Chapter 4, a system incurring setup times and no setup costs, when production is switched from one part type to the next, was studied. The

optimal transient solution was obtained by partitioning the production surplus space (inventory/backlog space) of the system into two mutually exclusive major regions. For initial production surplus levels in one region, the optimal transient solution was obtained by inspection. For initial production surplus levels in the other region, the optimal transient solution was first obtained by means of a novel algorithm. Then, based on insights from the numerical solution, the optimal transient solution was derived analytically. In Chapter 5, The results of Chapter 4 were extended to the more general case where both setup times and costs are incurred, when production is switched from one part type to the next. The complete optimal transient solution was obtained analytically as a feedback law.

In Chapter 6, a study of the stochastic system was presented. The problem was formulated as a stochastic optimal control problem, and necessary and sufficient optimal conditions were given as a system of nonlinear partial differential equations, known as the Hamilton-Jacobi-Bellman (HJB) optimality partial differential equations. A simple technique, due to Kushner and Dupuis (1992), was then adopted to convert the HJB optimality partial differential equations to a classical infinite horizon discrete time discrete state Markov Decision Process. The optimal production and setup controls were then obtained numerically by means of the classical value iteration algorithm of dynamic programming. The optimal production policy has been shown to be of the bang-bang type, where either a part type is produced at the maximum machine production rate or not produced at all. The optimal setup policy has been shown to have two different structures, depending on the parameters of the system: A Orthogonal Corridor Structure and a Parallel Corridor Structure. Given the computational difficulties of the exact technique, a simple heuristic policy (Hedging Corridor Policy), based on results from the deterministic

system, was then proposed to approximate the optimal policy. Simulation results showed that the heuristic policy gives a very good approximation for systems that are sufficiently reliable with similar characteristics of the produced part types.

7.2 Suggestions for Future Work

This research is only the start of work in a potentially fruitful arena of real-time production scheduling problems. There are many possible directions that can be taken to enhance and further develop solution methodologies for the complex dynamic setup problem beyond the work of this dissertation.

For the deterministic system, a natural extension is to consider a multi-part-type system. we already presented (Chapter 3) a nonlinear programming algorithm to obtain the optimal steady state solution of such a system. To obtain the optimal transient solution, a technique similar to the one developed in this dissertation can be adopted. First, the production surplus space has to be partitioned into two mutually exclusive major regions. One above the cyclic schedule and one below it. For initial production surplus levels above the cyclic schedule, optimal control techniques can be used to determine the optimal trajectory leading to the cyclic schedule from above. For initial surplus levels below the cyclic schedule, the trajectory leading to the cyclic schedule from below can be obtained by extending the algorithm, developed in Chapter 4, to handle multiple part types.

Another possible direction would be to analyze the setup problem for a deterministic system with multiple machines and multiple part types in which it is not possible to devote each machine to the production of a single part type.

The stochastic system can be extended to the three-part-type system. The latter is further complicated by the sequencing policy that has to be determined. The same numerical technique, developed in Chapter 6, can be applied to this case. By observing the numerical solution, simple heuristics, implementable in real-time, can be proposed. The Hedging Corridor Policy can also be tested on the three-part-type system. If it shows a good performance, then it can be extended to the multi-part-type system. The advantage of the Hedging Corridor Policy lies in its simplicity.

LIST OF REFERENCES

Akella, R., Choong, Y. and Gershwin, S. B. (1984), "Performance of Hierarchical Scheduling Policy," IEEE Transaction on Components, Hybrids, and Manufacturing Technology, CHMT-7(3), 225-240.

Akella, R. and Kumar, P. R. (1986), "Optimal Control of Production Rate in a Failure Prone Manufacturing System," IEEE Transaction on Automatic Control, AC-31(2), 116-126.

Arizono, I., Yokoi, S. and Ohta H. (1989), "The Effects of Varying Production Rates on Inventory Control," Journal of Operational Research Society, 40, 789-796.

Axsäter, S. (1985), "Performance Bounds for Lot Sizing Heuristics," Management Science, 31, 634-640.

Bai, S. X. and Burhanpurwala, J. H. (1993), "Real-Time Scheduling Algorithm for a Dynamic Setup Problem," SPIE'S International Symposium on Optics for Manufacturing and Advanced Automation. Boston.

Bazaraa, M. S. and Shetty, C. M. (1979), NonLinear Programming Theory and Algorithms. John Wiley & Sons, New York.

Bertsekas, D. P. (1987), Dynamic Programming. Deterministic and Stochastic Models. Prentice Hall, Inc., Englewood Cliffs, NJ.

Boctor, F. F. (1982), "The Two-Product, Single-Machine, Static-Demand, Infinite Horizon Lot Scheduling Problem," Management Science, 28, 798-807.

Bomberger, E. (1966), "A Dynamic Programming Approach to a Lot Scheduling Problem," Management Science, 12, 778-784.

Boukas, E. K. and Haurie, A. (1990), "Manufacturing Flow Control and Preventive Maintenance: A Stochastic Control Approach," IEEE transaction on Automatic Control, AC-35(9), 1024-1031.

Buzacott, J. A. and Ozkarahan, I. A. (1983), "One- and Two-Stage Scheduling of Two Products   with Distributed Inserted Idle Time: The Benefits of a

Controllable Production Rate," <u>Naval Research Logistics Quarterly</u>, 30, 675-696.

Carreno, J. J. (1990), "Economic Lot Scheduling for Multiple Products on Parallel Identical Processors," <u>Management Science</u>, 36, 348-358.

Connolly, S. (1992), "A Real-Time Policy for Performing Setup Changes in a Manufacturing System," Master's Thesis, MIT Operations Research Center, MIT Laboratory for Manufacturing and Productivity. Report LMP-92-0005, Cambridge.

Connolly, S., Dallery, Y. and Gershwin, S. B. (1992), "A Real-Time Policy for Performing Setup Changes in a Manufacturing System," <u>Proceedings of the 31st IEEE Conference on Decision and Control</u>, Tucson, Arizona, December 16-18.

Dobson, G. (1987), "The Economic Lot Scheduling Problem: Achieving Feasibility Using Time-Varying Lot Sizes," <u>Operations Research</u>, 35, 764-771.

Elmaghraby, S. E. (1978), "The Economic Lot Scheduling Problem (ELSP): Review and Extensions," <u>Management Science</u>, 24, 587-598.

Gallego, G. (1989), "Real-Time Scheduling of Several Products on a Single Facility With Setups and Backorders," Ph.D. Thesis, School of OR&IE, Cornell University, Ithaca, NY.

Gershwin, S. B. (1993), <u>Manufacturing Systems Engineering</u>. Prentice Hall, Inc., Englewood Cliffs, NJ.

Goyal, S. K. (1984), "Determination of Economic Production Quantities for a Two-Product Single Machine System," <u>International Journal of Production Research,</u> 22, 121-126.

Haurie, A. and L'Ecuyer P. (1986), "Approximation and Bounds in Discrete Event Dynamic Programming," <u>IEEE Transaction on Automatic Control</u>, AC-31(3), 227-235.

Hsu, W. L. (1983), "On the General Feasibility Test of Scheduling Lot Sizes for Several Products on One Machine," <u>Management Science</u>, 29, 93-105.

Hu, J. and Caramanis, M. (1995), "Dynamic Set-up Scheduling of Flexible Manufacturing Systems: Design and Stability of Near Optimal General Round Robin Policies," to appear on <u>Discrete Event Systems, IMA Volumes in Mathematics and its Applications Series</u>, Springer-Verlag, New York.

Jones, P. and Inman, R. (1989), "When is the Economic Lot Scheduling Problem easy?," <u>IIE Transactions</u>, 21, 11-20.

Kimemia, J. and Gershwin, S. B. (1983), " An Algorithm for the Computer Control of a flexible Manufacturing System," <u>IIE Transactions</u>, 15(4), 353-362.

Kushner, H. J. and Dupuis, P. G. (1993), <u>Numerical Methods for Stochastic Control Problems in Continuous Time</u>. Springer-Verlag, New York.

Kumar, P. R. and Seidman, T. I. (1990), "Dynamic Instabilities and Stabilization Methods in Distributed Real-Time Scheduling of Manufacturing Systems," <u>IEEE Transactions on Automatic Control</u>, 35, 289-338.

Lee, C. Y. and Surya, D. (1989), "Economic Lot Scheduling for Two-Product Problem," <u>IIE Transactions</u>, 21, 162-169.

Lou, S., Sethi, S. and Sorger, G. (1991), "Analysis of a Class of Real-Time Multiproduct Lot Scheduling Policies," <u>IEEE Transactions on Automatic Control</u>, AC-36(2), 243-248.

Lou, S., Sethi, S. and Sorger, G. (1992), "Stability of Real-Time Lot Scheduling Policies for an Unreliable Machine," <u>IEEE Transactions on Automatic Control</u>, AC-37(12), 1966-1970.

Maxwell, W. L. and Singh, H. (1983), "The Effect of Restricting Cycle Times in The Economic Lot Scheduling Problem," <u>IIE Transactions</u>, 15, 235-241.

Moon, I., Gallego, G. and Simchi-Levi, D. (1991), "Controllable Production Rates in a Family Context," <u>International Journal of Production Research</u>, 29, 2459-2470.

Newson, E. F. P. (1975a), "Multi-Item Lot Size Scheduling by Heuristic, Part I: With Fixed Resources," <u>Management Science</u>, 21-10, 1186-1193.

Newson, E. F. P. (1975b), "Multi-Item Lot Size Scheduling by Heuristic, Part II: With Variable Resources," <u>Management Science</u>, 21-10, 1193-1203.

Olsder, G. J. and Suri, R. (1980), "Time Optimal of Parts-Routing in a Manufacturing System with Failure Prone Machines," <u>Proceedings of the 19th IEEE Conference on Decision and Control</u>, Albuquerque, NM, 722-727.

Perkins, J. and Kumar, P. R. (1989), "Stable, Distributed, Real-Time Scheduling of Flexible Manufacturing/Assembly/Disassembly Systems," <u>IEEE Transactions on Automatic Control</u>, 34, 139-148.

Rishel, R. (1975), "Dynamic Programming and Minimum Principles for Systems with Jump Markov Disturbances," <u>SIAM Journal On Control</u>, 113(2), 338-371.

Roundy, R. (1985), "98%-Effective Integer-Ratio Lot Sizing for One-Warehouse Multi-Retailer Systems," <u>Management Science</u>, 31, 1416-1430.

Sharifnia, A., Caramanis, M. and Gershwin S. B. (1991), "Dynamic Setup Scheduling and Flow Control in Manufacturing Systems," <u>Discrete Event Dynamic Systems: Theory and Applications</u>, 1, 149-175.

Silver, E. (1990), "Deliberately Slowing Down Output in a Family Production Context," <u>International Journal of Production Research</u>, 28, 17-27.

Sivazlian, B. D. and Stanfel, L. E. (1975), <u>Analysis of Systems in Operations Research</u>. Prentice Hall, Inc., Englewood Cliffs, NJ.

Srivatsan, N. and Gershwin, S. B. (1990), "Selection of Setup Times in a Hierarchically Controlled Manufacturing System," <u>Proceedings of the 29th IEEE Conference on Decision and Control</u>, Honolulu, HI, December.

Tsitsiklis, J. N. (1982), "Characterization of Optimal Policies in a Dynamic Routing Problem," <u>M.I.T. Laboratory for Information and Decision Systems</u> Report No. LIDS-r-1178, Cambridge.

Wagner, H. M. and Whitin, T. M. (1958), "Dynamic Version of the Economic Lot Size Model," <u>Management Science</u>, 5(1), 56-63.

Zipkin, P. (1991), "Computing Optimal Lot Sizes in the Economic Lot Scheduling Problem," <u>Operations Research</u>, 39, 53-63.

BIOGRAPHICAL SKETCH

Mohsen El Hafsi was born January 7, 1963, in Tunis, Tunisia. He received primary and secondary education in Tunis, Tunisia. In June of 1982, he graduated from high school and received his baccalaureate diploma, with high honors, from Lycée Bab El Khadra de Tunis. He attended the Ecole Nationale d'Ingénieurs de Tunis, Tunisia, from 1982 to 1988, receiving the Ingénieur Principal's (Qualified Engineer) degree in industrial engineering, with high honors, in July 1988. From October 1988 to August 1990, he worked as a research engineer in the Institut Regional des Sciences Informatiques et des Télécommunications. In August 1990, he was awarded a scholarship from the Tunisian government to pursue his graduate studies in the United States of America. In January 1991, he joined the Industrial and Systems Engineering Department at the University of Florida to pursue a Master of Science and Doctor of Philosophy degrees.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
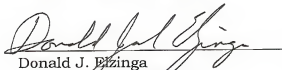
Boghos D. Sivazlian, Chairman
Professor of Industrial and
Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sherman X. Bai, Cochairman
Assistant Professor of Industrial
and Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
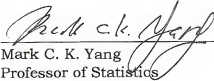
Donald J. Elzinga
Professor of Industrial and
Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Chung-Yee Lee
Professor of Industrial and
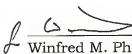 Systems Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mark C. K. Yang
Professor of Statistics

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1995

Winfred M. Phillips
Dean, College of Engineering

Karen A. Holbrook
Dean, Graduate School